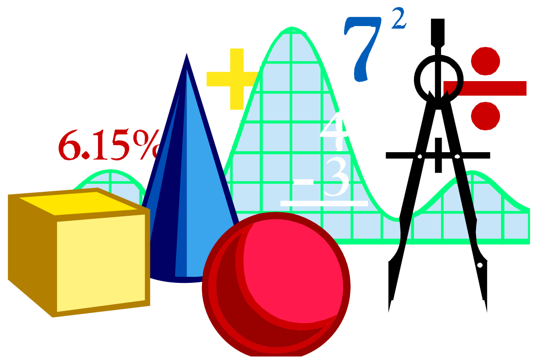


STATISTICAL LITERACY 2021B

Seeing the Story Behind the Statistics

Milo Schield

STATISTICAL LITERACY 2021B:

Seeing the Story behind the Statistics

Eighth Edition

Printed in the U.S. by Instant Publisher

ISBN: 978-1-61422-671-0

File name: 2021B-SL4DM-V1F.docx Size: 172,280 kb Save: 12/20/2021

499 pages, 7901 paragraphs, 15,275 lines, 143,172 words, 732,405 characters.

228 figures, 129 tables, 34 stories, 11 equations,
210 glossary terms, 420 index entries,

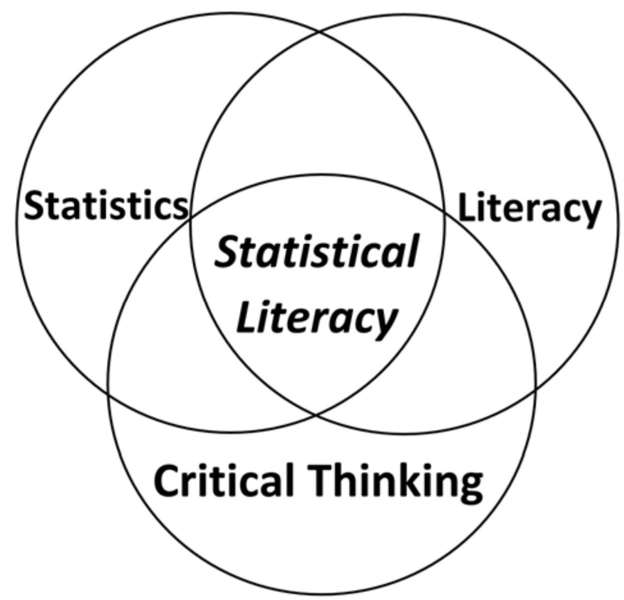


TABLE OF CONTENTS

Chapter 1. The story behind the statistics	**7**
Chapter 2. Comparisons and CARE Solutions	**59**
Chapter 3. Understanding measurements	**114**
Chapter 4. Describing ratios: Percent and Percentage	**160**
Chapter 5. Describing Ratios: Rate, Chance and Ratio	**196**
Chapter 6. Comparing Ratios using Named-Ratios or Likely	**239**
Chapter 7. Interpreting Confusing Ratios	**271**
Chapter 8. Randomness	**317**
Key Word List by Chapter	**342**
Table of Figures	**344**
Table of Tables	**350**
Table of Stories	**353**
Table of Equations	**354**
Glossary	**355**
Index	**369**
Review of Named Ratios	**375**

Typography: Bold indicates a term that is being defined. Any word or phrase in bold should be found in the Glossary. Italics indicate an important concept, point or claim. In part-whole ratios, underscoring 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 decision makers
make better decisions using statistics as evidence

[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, Relative Risk and the Cornfield conditions, and a Fellow and President of the American Statistical Society—used observational statistics to argue that the association between smoking and cancer was "so great" that it cannot be due to any known confounder. Dr. Dennis Haack's 1979 textbook, "*Statistical Literacy*" was the first book to use 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 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 social products, the results of people's efforts." He argued that understanding this fact is most essential in evaluating statistics as evidence in arguments. He also noted that "Statistics can become weapons in political struggles over social problems and social policy."

Preface

Table of Contents:

Preface 1
Students: See the story behind the statistics 1
Secondary Goal 2
Teachers: Statistical literacy solves a problem 2
Teachers: Is this Statistics? 4
Audience for this book 4
Design and Use of this book 4
Statistical Literacy and GAISE 4
Ways to use this textbook 5
Recommended Supplements 5
Statistical Literacy: Descriptions and Comments 6

Preface

Students: See the story behind the statistics

You may be very interested in social issues such as social justice or individual rights. Most of these involve statistics: social statistics. Figuring out whether the statistics provide strong evidence is not easy. This course is dedicated to helping you read, interpret and evaluate social statistics. Social statistics are very different from numbers. Usually there is a story behind the statistics. That story is important; it deserves to be told, studied and understood.

Social statistics may involve sensitive subjects. If you are an adult and some social statistics trigger strong feelings, then your job is to deal with your feelings. If you are having difficulties, talk to your instructor or counselor. Ultimately, you are responsible for controlling your mind. If you can't, then you are vulnerable to being controlled by others who may not have your best interests at heart.

Not all statistics are your friends. Some statistics are false, while others are ambiguous, ill-defined, misleading or spurious. As Mark Twain said, "There are three kinds of lies: lies, damned lies and statistics."

As a reader you have three choices: (1) be naïve: treat every statistic as a fact, (2) be a cynic: ignore every statistic since it might be an error or an opportunistic misrepresentation, or (3) be a critical thinker: learn how to distinguish good and bad statistics. You may get overwhelmed or feel trapped when you see a statistic. If you accept the statistic as strong evidence, you may be fooled. If you reject a statistic as evidence, you may be wrong and make a bad choice.

The primary goal of statistical literacy is to help you evaluate the credibility of a statistic (not just the source) — to go beyond reading statistics in the everyday news 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 untangle an argument. It takes more training and practice to untangle an argument that uses statistics as evidence. A goal of this book is to give you that training and practice.

Statistical literacy is the ability to read and interpret statistics in the everyday media. Statistical literacy is critical thinking about statistics as evidence in arguments.

Statistical literacy is closer to critical thinking than to mathematics. Mathematics is deduction: true or false, right or wrong. Critical thinking involves induction: shades of grey, strength of evidence supporting a conclusion. Think of statistical literacy as quantitative rhetoric[2].

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 social statistics—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 a new 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 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 read and communicate arithmetic comparisons, ratios (using percentage, rate and change grammar) and comparisons of ratios (using likely grammar).

Teachers: Statistical literacy solves a problem

Traditional introductory statistics has a problem. Most students see less value in statistics after that course that they did before taking the course.

[2] Schmit, John (2010) Teaching Statistical Literacy as a Quantitative Rhetoric Course. *Proceedings of Section on Statistical Education*. www.statlit.org/pdf/2010SchmitASA.pdf

The traditional inference course is designed for Producers: those who major or minor in statistics and those who may conduct or read statistical studies.

Most college graduates are Consumers. They will never conduct or read the results of a study. They will see statistics in the everyday media that they need to read and interpret. Many will make de7cisions based on statistics.

Students see value in statistical literacy. Most agree that statistical literacy should be required for graduation by all students.

Statistical literacy is different:

- Different audience: Statistical literacy is for those who need to read and interpret the statistics they encounter in the everyday media.
- Different content: Traditional statistics focuses on randomness; statistical literacy focuses more on confounding.
- Different focus: The world has changed. In the past, small surveys and clinical trials were more common. Today, large surveys, big data, quasi-experiments, and observational studies are much more common.
- Different approach: Statistical inference is deductive: if the premises are true, then the conclusion must be true. In random sampling, 95% of all the 95% confidence intervals must include the population parameter. Statistical literacy is inductive: how strong is the evidence in supporting a disputable claim or conclusion?
- Different thinking: Traditional statistics involves mathematical thinking: Statistical literacy involves hypothetical thinking. What should have been controlled for? How could things have been defined, counted and measured differently? What was omitted or overlooked? What difference would these differences make?
- Different mission: The mission of statistical literacy is to improve critical thinking about statistics as evidence in arguments. With the advent of big data, statistics are everywhere. Most social arguments involve social statistics as evidence. Students need to read and evaluate these statistics.

Students see that studying statistical literacy helps them deal with today's big issues. Statistical literacy helps them understand and evaluate the statistics presented in statements, tables and graphs. Students learn how to describe and compare these statistics using ordinary English.

As one college senior said, "Statistical literacy is the hardest course I've taken, but I think it will be the most valuable."

If you teach the traditional introductory statistics course, but want more of your students to see value in studying statistics, give statistical literacy a try.

Teachers: Is this Statistics?

How can statistical literacy be statistics if statistical literacy has less than a 30% overlap with the traditional statistical-inference course?

Statistics studies variation. Variation is classified as random or systematic. Traditionally, the first statistics course focuses on random; the second focuses on systematic. The problem is that most students never take the second.

Statistical literacy studies variation – systematic and random – with a stronger emphasis on systematic. It includes topics from both the first and second statistics courses, and from epidemiology.

Audience for this book

This book is designed for consumers and decision makers. Decision makers include those who access, assemble, prepare and disseminate the statistics.

Decision makers make decisions involving large structured groups; consumers make decisions involving themselves or small informal groups. At college, decision makers are mainly those in non-science quantitative majors (the professions and the social sciences) or in political science, communications or journalism. Those in non-quantitative majors are typically consumers.

Design and Use of this book

By limiting the topics, this book allows instructors free time to experiment and do what they want: analyze news stories, investigate journal articles or run a project. Learning to think critically about statistics takes time and lots of practice. This book allows instructors the time to help students reach that goal.

This book is suitable for two-year and four-year colleges and in schools as an alternative to AP statistics. It has been used in an on-line class. This text can be used in a bridging course for those who will take statistical inference.

This book can be used as a companion text or as a text for a full course. A Moodle test-bank is available on Classroom Revolution.

Statistical Literacy and GAISE

This textbook is compliant with the first three recommendations of the ASA GAISE 2005 College Report: introductory statistics courses should "strive to (1) emphasize statistical literacy and develop statistical thinking, (2) use real data and (3) stress conceptual understanding rather than mere knowledge of procedures. This report suggested assessing statistical literacy by students "interpreting or critiquing articles in the news and graphs in media." *This textbook is compliant with the GAISE 2016* recommendation for more focus on multivariate thinking and confounding.[3]

[3] Schield, Milo (2017). GAISE 2016 Promotes Statistical Literacy.

Ways to use this textbook

This textbook has less than a 30% overlap with the topics in a traditional statistical inference course. As such it provides a unique alternative to traditional statistics.

As a separate course, start by using just the first three chapters. They have the main ideas. This allows plenty of time to focus on reading news stories. Plan A (Inference emphasis): add the last two chapters. Plan B (confounder emphasis): Cover Ch 1-4 and 7; skip 5, 6 and 8. Add skipped chapters later.

In a traditional statistics course, use the last three weeks to introduce confounding in chapters three and seven and show how controlling for confounding can influence statistical significance in observational studies as shown in chapter eight.

For those who are averse to teaching English grammar and strength of evidence, focus on chapters three, seven and eight where problems have right-wrong answers.

Here are some of the unique topics in this textbook:

Ch 1: Statistics are Numbers in context. Take CARE: four kinds of influence.

Ch 2: Comparison: Arithmetic and Grammar (includes 'times less') [ESL]
Study designs including quasi-experiments (along with letter grades)
Randomness: Law of Very Large Numbers

Ch 3: Standardization of measures (algebraic and graphical)
Cornfield's conditions for nullifying or reversing an association

Ch 4: Describing part-whole ratios using percent and percentage grammar [ESL]
Reading 100% tables, half tables and two-way half tables.

Ch 5: Describing part-whole ratios using rate and chance grammar [ESL]

Ch 6: Comparing part-whole ratios using percentage and likely grammar [ESL]

Ch 7: Diabolical denominators. Interpreting part-whole ratios: medical tests.
Standardization of percentages. Cornfield conditions.

Ch 8: Statistical significance influenced by confounding, assembly and bias.
Sampling designs (along with letter grades)

Recommended Supplements

- 2020: *Calling Bullshit: The Art of Skepticism …* by Bergstrom and West
- 2019: *Critical Statistics* by Robert de Vries
- 2019: *Guided Worksheets for Thinking Quantitatively* by Eric Gaze
- 2015: *Thinking Quantitatively: Communicating with Numbers*, Eric Gaze
- 2012: *Just Plain Data Analysis* by Gary Klass
- 2010: *Case Studies for Quantitative Reasoning* by Bernie Madison, et al.

Statistical Literacy: Descriptions and Comments

"What is statistical literacy? What every educated person should know."[4]
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."[5] 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."[5] Lynn Steen, past-President of the MAA.

"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."[5] 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 (2003)

"From my perspective, this teaching of causal inference is the most interesting topic today in statistical education, certainly so at the undergraduate level" [5]
Dr. Donald Rubin, Professor of Statistics, Harvard University.

"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.

I hope that...statistical literacy will...rise to the top of your advocacy list. Ruth Carver, 2012 ATOMPAV Presidential Address[6]

"Statistical literacy should be part of every citizen's tool kit" Nicholas Kristof[7]

Statistical literacy has risen to the top of my advocacy list, right alongside numeracy, and perhaps even ahead of "algebra for all." Michael Shaughnessy[8], Past-President of the NCTM. USCOTS 2015 Lifetime Achievement award.

---

[4] http://www.statlit.org/Moore.htm
[5] Private communication (2001)
[6] http://www.statlit.org/pdf/2012-Carver-Presidents-Message-Statistical-Literacy.pdf
[7] http://www.statlit.org/pdf/2015-Kristof-NY-Times-0425.pdf
[8] http://www.statlit.org/pdf/2010Shaughnessy-StatisticsForAll-NCTM.pdf

Chapter 1. The story behind the statistics
TABLE OF CONTENTS

Learning Outcomes	7
Introduction	8
Statistics are Numbers in Context	10
Statistics in Arguments	12
What kinds of things influence statistics?	14
Summary	15
Critical Thinking Issues	16
Association	16
Causation	17
ABC Keywords for Association vs Causation	19
Disparity vs. Discrimination	21
Association is Not Causation	22
Strength of evidence	24
Statistical Literacy Issues	26
C = CONFOUNDING	27
A = ASSEMBLY	32
R = RANDOMNESS	37
E = ERRORS	42
Putting It All Together: Analyzing Stories	47
Chapter Summary	50
Optional	51
Arguments	51
Between words	53
Mechanism	54
Separate Statistical Literacy from Critical Thinking	56
Seeing pure randomness	56

Learning Outcomes

Critical thinking outcomes:

- Distinguish association from causation using an ABC classification.

Statistical literacy outcomes:

- Learn that statistics are numbers in context: they can be influenced.
- Learn that statistics are socially constructed by people with motives and values
- Learn that "take care" is the best advice when dealing with statistics.
- Learn how all influences on statistics can be grouped into four categories.
- Learn CARE categories: Confounding, Assembly, Randomness and Error
- See how each influence can change the size and direction of an association.
- Use Take CARE to analyze statistics in the everyday media

Introduction

The goal of statistical literacy is for you to be able to read, interpret and evaluate the social statistics you encounter in the everyday media.

Consider two news stories published in the same month:

- An article in the British Medical Journal claims "There is overwhelming evidence that excessive consumption [of salt] causes high blood pressure... "
- An article in the Journal of the American Medical Association concludes that "Dietary salt intake has little effect on blood pressure..."

How can these opposite claims be published at the same time? To address that question, you need statistical literacy, the subject of this text. **Statistical literacy** studies statistics in everyday usage. Statistically literate people think critically about the statistics behind claims like these.

Let's see what statistical literacy means in thinking critically about this story:

Story 1: Fruits, Veggies Cut Risk of Breast cancer

Fruits, Veggies Cut Breast Cancer Risk
Four or more daily servings reduces chances of disease by half

FRIDAY, Oct. 31 (HealthDayNews) -- A diet rich in fruits and veggies can help protect against breast cancer.

A study by Oregon Health and Science University researchers found women who eat at least four servings of fruits and vegetables have a 50 percent lower risk of breast cancer than women who consume no more than two such servings each day.

They reached that conclusion after examining the diets of 378 women with breast cancer and the diets of 1,070 cancer-free women. All the women, living in Shanghai, China, filled out questionnaires that asked about their intake of 108 individual food items, fried and restaurant food, dietary changes, and the use of nutrient supplements and Chinese herbal medicines.

First we must decide what the point of the story is. You can see it from the first sentence: *Eating lots of fruits and veggies can help protect against breast cancer*. The article would have had a different point if it had said *Breast cancer was less prevalent among women who ate lots of fruits and veggies than among those who ate little or none*. The latter point reports a factual matter; the point actually made claims that there's a cause. Causation is also implied by the action words used in the story: *cut* and *reduces.* One aspect of statistical literacy is distinguishing claims of association from claims of causation.

Another aspect of statistical literacy is thinking about the statistics themselves. Statistics are numbers describing data. But they're not pure numbers; they're numbers that measure or count real things. Therefore they can be shaped by

Chapter 2. Comparisons and CARE Solutions

TABLE OF CONTENTS

Learning Outcomes	59
Review	60
Association	60
Two-Group Comparisons	60
Two-Factor Co-variation	68
Confounder Solutions	69
Study Design	70
Experiments vs. Observational Studies	70
Effect Size	79
Selection and Matching	82
Assembly Solutions	85
Assembly in Definitions	86
Assembly in Comparisons	87
Randomness Solutions	91
Extremes and Coincidence	91
Big Data and the Law of Very Large Numbers	92
Small Samples, Margin of Error and Statistical Significance	94
Error (Bias) Solutions	98
Three Kinds of bias	98
Minimizing bias	100
Analyzing Stories	101
Summary	103
Optional	105

Learning Outcomes

Recognize associations as possible indications of causation

- Analyze associations: two-group vs. two-factor
- Study two-group comparisons: non-arithmetic (ordinal) vs. arithmetic.
- Distinguish four kinds of arithmetic associations and exceptions.

Recognize three techniques that deal with confounding

- Distinguish different kinds of experiments and observational studies
- Recognize how different study designs control different confounders.
- Know that a controlled study is "any study involving a control group"
- Distinguish 'control of' and 'control for'.

Study techniques that deal with randomness and bias

- Use overlapping confidence intervals to test for statistical significance
- Recognize bias-control methods: placebo, single and double blind.

Review

Chapter 1 focused on problems: the influences on a statistic. It contained a lot of new material. Here are a few of the highlights.

- showed that statistics are typically used as evidence in arguments.
- introduced statistics as numbers in context, so they can be influenced.
- admonished students to "Take Care" when dealing with statistics.
- used CARE to indicate the four kinds of influence on a statistic.

Review the chapter summary at the end of chapter 1. Review each figure. Identify what is being stated. Understand why the information in the figure is important.

Become familiar with the key words and phrases in the new word list. Becoming statistically literate is like learning a new language. It takes time and repetition.

This chapter introduces some of the solutions to some of the problems in chapter 1. It introduces some things to look for in analyzing the use of statistics as evidence.

Association

As noted in chapter 1, there are two kinds of association: two group comparisons and two-factor co-variation. Both are examined in more detail.

Two-Group Comparisons

The most common form of an association is a two-group comparison. There are three types of two group comparisons[72]: raw comparisons, ordered comparisons and arithmetic comparisons. **Raw comparisons** just present the values to be compared. Raw comparisons provide the underlying data, but the user must figure out the order and the size of the comparison. Ordered comparison provide order, but lack precision. Arithmetic comparison provide order and precision.

Ordered Comparisons

Ordered comparisons compare two values without mentioning their size or the size of the comparison. "Seniors are older than Juniors," Ordered comparisons are common in headlines and advertisements.

Table 3 presents counts of accidental deaths.

[72] Others. Ordered raw: Male height (69") is greater than female height (65"). Arithmetic raw: Male height (69") is 4" more than female height (65"). Raw obscures the comparison.

To see this, study the above worksheet[151]. There are 28 people at a table: seven on each side. That gives 49 combinations (7*7) between adjacent sides (top and right; top and left; bottom and right; bottom and left) and opposite sides (top-bottom and left-right). There are four sets of adjacent sides and two sets of opposite sides. This gives a total of 294 pairs given six sets at 49 per set. Finally, there are 21 pairs (6+5+4+3+2+1) for two matches within each of the four sides for a total of 84 pairs (4*21) giving a grand total of 378 pairs (294+84).[123]

With 378 possible pairs hidden inside these 28 people, we can expect one match given one chance in 360 of a match. At least one match is more likely than not.

[151] Assumes 30 days for each month. www.statlit.org/Excel/2012Schield-Bday.xls

Chapter 3. Understanding measurements

TABLE OF CONTENTS

Learning Outcomes 114
Review 115
Data and Data Distributions 115
 Continuous Data 116
Measures of Location 118
 Ranks 118
 Percentiles 119
Measures of Center 122
 Mean, Median and Mode 122
 Measures of Center vs. Skew 124
 Describing and comparing centers of measures 125
 From Association to Causation 127
Assembly on Measures of Center 130
 Assembly: Defining Groups and Group Averages 130
 Assembly: Forming Groups by Choosing a Cut Point 130
 Assembly: Choosing a Measure of Center 131
Controlling for Two-Group Confounding 132
 Forming and Comparing Per Ratios 134
 Untangling Confounding of Averages using Selection 136
 Standardizing Averages Arithmetically: Mix Matching 138
 Standardizing averages graphically 142
 Percentage explained by a confounder 150
 The Cornfield Conditions 152
 Controlling for Confounding involving Co-Variation 153
 Spread 156
Summary 157
Optional 157
 Two-Group Comparisons of Measures (Detail) 157

Learning Outcomes

Understand ranks and percentiles.

Recognize the order of mean, median and mode in skewed distributions.

- Compare means, medians or modes using ordinary English.
- Distinguish percent, "percentage points" and "percentile points".
- See how assembly can influence the size of a comparison.

See how shift from totals to ratios can change size & direction of a comparison.

- Compute weighted average from subgroup averages.

Calculate the influence of a binary confounder on a comparison of two means.

- Standardize using arithmetic mix-matching or a graphical technique
- Use Cornfield conditions to see if nullification or reversal is possible.

c. Subject is the group. The measure is inside the verb. The determiners are in trailing prepositional phrases

Means: Adults in 2000 are 10 cm taller on average than in 1920.
Modes Adults in 1920 are typically 10 cm taller than in 1920.
Median: Do not use

===

#3: Compare two measures (measure1 and measure2) involving one group at one time and place. Assume all involve US families in 2018.

a. Subject is measure1. Measure2 is in the predicate. Group possessing the measure is in a prepositional phrase.
Long: The average income of families was 25% greater than the median income of families.
Short: Average family income was 25% greater than median income

*b. Subject is a phrase involving the group (a possessive adjective) and measure1 (a noun). Sentence predicate is a phrase involving a possessive adjective (groups') and a noun (measure2)**
Long: Families' mean income was 25% more than families' median income.
Medium: Families' mean income was 25% more than *their* median income.

c. Subject is the group; measure1 is in the predicate. Comparison is in relative clause modifying measure1.
Short: Families have an average income that is 25% greater than their median income.

Discussion:
It is tempting to say "the average adult male" has a height of 5' 10" instead of saying "The average height of adult males is 5'10". . That usage can be problematic with qualities: The average adult has one testicle and one breast. It is somewhat problematic with discrete quantities: "The average person has less than two legs." Although it is not problematic when describing a continuous quantity, it can be problematic in comparing such a measure for two groups: "The average city in California is bigger than that in Iowa". We don't know the measure. Is it population or area? We don't know. It seems better to connect the measure of center (average, median or modal) with the measure it modifies in a single adjective-noun phrase.

Chapter 4. Describing ratios: Percent and Percentage

TABLE OF CONTENTS

Learning Outcomes	160
Review	161
Part-Whole Ratios	161
Part-Whole Ratios	163
Pie Charts of Counts and Ratios: Good and Bad	163
Percent Grammar using Tables of Counts	165
Percent Grammar Statements	166
Percent Grammar: Forming Statements	166
Common Activities involving Count Tables	168
100% Ratio Tables: Three Kinds	169
Decoding Problems	174
Percentage Grammar	176
Reading percentages in 100% tables	178
Count tables	179
Sports and Portion Grammar	181
Who-Who or Who-What Ambiguity	181
Distribution Grammar	182
Half tables of percentages	183
One-way half-tables of percentages	183
Two-way half-tables of percentages	186
Tables and graphs with missing margins	188
Conclusion	192
Optional	193
Convert statements from percentage to percent grammar	194
Convert statements from percent to percentage grammar	194
Convert questions between Percent and Percentage grammar.	195

Learning Outcomes

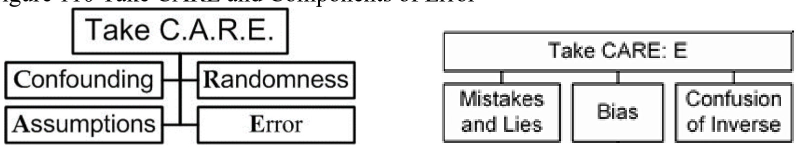
Learning outcomes include:

- Identify part and whole in named ratio grammar statements and questions.
- Learn named-ratio grammars: percent and percentage grammar
- Distinguish percent and percentage grammar.
- Translate part-whole statements between percent and percentage grammar.
- Describe percentages ratios in 100% tables and in half-tables.
- Describe percentages in tables with missing margins.
- Recognize the importance of selection and the confusion of the inverse.

Review

In chapter 1, we noted that statistics – unlike numbers – can be influenced. These influences were classified into four groups as the left side of Figure 110 reminds us. The right side points out the confusion of the inverse as a source of Error.

Figure 110 Take CARE and Components of Error



Confounding provides an alternate explanation for an association. Assembly includes how things are created, selected, defined, measured and presented.

In chapter 2, we studied four ways to form a two-group comparison. Of these, three involved ratios. Confounders can be controlled physically or mentally as shown on left side of and right side of Figure 111 respectively.

Figure 111 Controlling Confounders: Control Of and Control For

CONTROL OF CONFOUNDERS			
Physical Control (Grade = Quality)			
Experiment		Observational Study	
A+	Scientific	C	Longitudinal
A-	Random Assign	D	Cross-sectional
B	Quasi-Exper	F	Anecdotal story

CONTROLLING FOR CONFOUNDERS			
Take into account (mental)			
Can do by hand		*Calculator/Computer*	
1	Select/Stratify	4	Linear Regression
2	Form Ratios	5	Logistic Regression
3	Standardize	6	Multivariate Regress

Chapter 3 studied measurements. We studied three easy ways to control for measurement-related confounders: selection, ratios and standardization. The ratios were averages: sum of measures divided by their count.

Ratios are a natural way to control for the influence of a proportionally-related factors such as the size of a group. That may be why ratios are the most common statistic in everyday life. Ratios are the workhorse in the statistical literacy stable.

This chapter studies ratios of counts where the numerator is typically a part of the denominator. Mathematically, this is no big deal. But describing these ratios is a very big deal. It seems fairly simple, but it is not.

Part-Whole Ratios

To be statistically literate, one must be able to read and interpret part-whole ratios in text, tables and graphs.

The simplest way involves prepositions. Third-graders are introduced to count-based ratios using common prepositions such as *out of* (four out of five doctors recommend Crest) or just *of* (four *of* the six slices of pizza had peperroni). Adults may use a less-common preposition: *per*. Less than one infant *per* hundred dies from a Covid19 infection.

Prepositions are words that identify specific relationships: spatial (in, out), temporal (before, after), logical (if, then) or quantitative (out of, per).

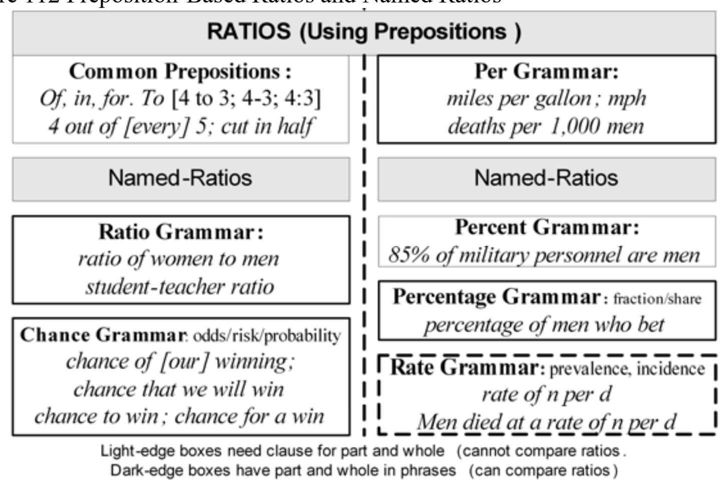
Part-whole ratios can be expressed using just prepositions. But it is difficult to form comparison of these ratios. By giving some of these ratios names, we can describe and compare ratios more effectively.

In English, there are certain nouns that indicate a particular type of ratio and do so with a unique grammar. In this book, these are called named ratios.[192]

Named ratios are words that indicate the presence of a ratio: nouns such as *rate*, *percentage* or *chance*, or adverbs such as *likely* or *prevalent* in comparisons.

Named ratios can be classifed into families that share a common grammar. The five main families are named the *percent*, *percentage*, *rate*, *chance* and *ratio* families. These five-named ratio families are grouped in two columns in Figure 112 depending on whether they involve 'per'.

Figure 112 Preposition-Based Ratios and Named Ratios



The 'per' statistics are arguably the most common ratios.[193] These appear in *percent*, *percentage* and *rate* grammar. Of these three named-ratio families, *percent* and *percentage* are the two that are the most error-prone. So, this chapter studies *percent* and *percentage* grammar. The next chapter studies the other three.

Treat each named ratio grammar as a new language. You will discover how small changes in syntax (grammar) can produce big differences in semantics (meaning).

Percent signifies per 100. Per 100 can be written in two ways: *percent* (in the US) and *per cent* (in the UK). This text uses the US style.

[192] www.statlit.org/pdf/2000SchieldASA.pdf
[193] Allan Tarp (2000) introduced the 'per' ratios. ICME-9. Tokyo.

Convert questions between Percent and Percentage grammar.

When used in a question, the keyword *percentage* is used with percent grammar. What percentage of men who run are seniors? This questions is really percent grammar. The whole and part are in the subject and predicate respectively, so seniors is the part. We don't ask, "What percent….?" That would be like asking "What pounds are you?" instead of "What weight are you?"

The keyword percentage can be used in a question with percentage grammar. What is the percentage of men who run? The presence of *the* before *percentage* indicates percentage grammar, so *runner* is the part.

1. To convert questions from percentage to percent grammar simply drop "is the" in "What is the percentage" and convert the relative clause to a main verb. For example, convert "What is the percentage of men who smoke?" to "What percentage of men are smokers?"
2. Converting questions from percent to percentage grammar is straightforward when there is no relative clauses. Change "What percentage" to "What is the percentage". Change the predicate into a relative clause. For example, convert "What percentage of men are smokers" to "What is the percentage of men who are smokers?"
3. In percent-grammar questions with one or two relative clauses, review the options given in converting percent statements to percentage. Convert "What percentage of <whole1> who are <whole2) are <part1> who are <part2>? Form the complete part as a single phrase without a trailing relative clause by moving the trailing adjective to a leading adjective. "Among <whole1> who are <whole2>, what is the percentage who are <part2><part1>?"

Convert these questions from percent to percentage grammar:
1. What percentage of college students are smokers?
2. What percentage of college students who are 18-25 are women?
3. What percentage of college students are women who smoke?
4. What percentage of college students who are 18-25 are women who smoke?

Warning: Students may get overloaded with the details of percentage grammar. To skip these details, use these two rules for percentage statements: "The percentage who" always introduces a part. "The percentage of" always introduces a part unless followed by a relative pronoun such as *who*, *what* or *that*.

Chapter 5. Describing Ratios: Rate, Chance and Ratio

TABLE OF CONTENTS

Learning Objectives	196
Review	198
Rate Grammar	199
Introducing rates	199
Phrase rate grammar	203
Clause rate-grammar	206
Chance Grammar	208
Ratio Grammar	211
Conclusion:	213
Optional	214
Social Statistics in Graphs and Tables	215
Household (HH) Statistics	215
Income Earned and Taxes Paid	216
Income Quintile Statistics	217
Income Mobility Statistics	217
Loan Interest-Principle Ratios	218
1860 U.S. Race Census by Geographical Region	219
Social Statistics Involving Children by Race	220
Count Statistics: Murder Victims	221
Law Officers Killed	221
Crime Statistics	222
Income Statistics	225
Clinical Trials	226
Longitudinal Graphs	228
Cross-Sectional Graphs	234

Learning Objectives

1. Rate grammar is most common in dealing with very small ratios.
2. The rate family includes prevalence and incidence
3. The keyword *rate* is equivocal. It have four forms. It can be a frequency (# events per unit time), a growth rate (% increase per unit time), a prevalence (# of unemployed per 100 members of the civilian labor force) or an incidence (# of births per year per 100,000 women ages 15-44).
4. Rate grammar has both a phrase form and a clause form.
5. The *chance* family of grammars includes odds, risk, likelihood and probability.
6. Chance grammars has both a phrase form and a clause form.
7. Chance grammars are common but may often leave the whole unstated.
8. The ratio family can easily be involved in the confusion of the inverse.

9. Graphs: recall that a quintile is one-fifth, quartile is fourth and decile is a tenth.
10. Ratios are easily influenced by the choice of the denominator.

Review

In chapter 1, we noted that statistics are numbers in context. Statistics can be influenced. The influences were classified into four groups: Confounding, Assembly, Randomness and Error/Bias. "Take CARE" reminds us of these.

Chapter 3 identified three easy ways to mentally control for confounding: the influence of a related factor. See Figure 135.

Figure 135: Controlling confounding: Control of vs control for

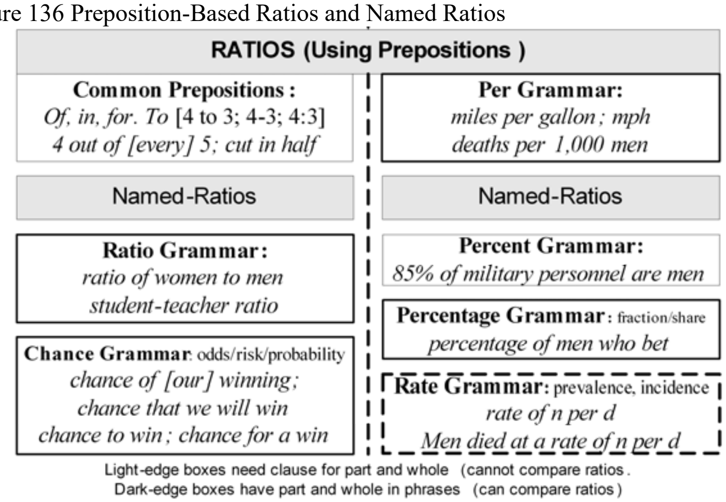
CONFOUNDING

Control of	Control for
Experiments	Selection, ratios
Observ. studies	Standardize

Comparisons, selection, ratios and standardization all involve assembly. Compared to what? Selected what? Out of what? Standardized by what?

Ratios are the workhorse in the statistical literacy stable. Chapters 4 introduced part-whole ratios using *percent* and *percentage* grammar. *Among* always indicated a whole, while *of* sometimes introduced a whole. Figure 136 classifies the ratios by the prepositions involved and introduces the five named ratios.

Figure 136 Preposition-Based Ratios and Named Ratios



Percent and percentage grammar are extremely susceptible to the confusion of the inverse. Describing percentages in tables and graphs can be tricky.

This chapter studies the three remaining named ratios: rate, chance and ratio.

Chapter 6. Comparing Ratios using Named-Ratios or Likely

TABLE OF CONTENTS

Learning Outcomes	239
Review	240
Introduction to Comparisons of Ratios	240
Attributable Comparisons	241
Percentage Attributable (Rate based)	242
Cases (Deaths) Attributable (Rate based)	243
Comparing Ratios: Common-vs-Distinct Part	245
Comparing Ratios Using Named Ratios	247
Reading ratio comparisons using named ratios	247
Constructing named-ratio comparisons	249
Generating short form comparisons	254
Spotting errors in comparisons	254
Comparing part-whole ratios presented in graphs:	255
Comparing ratios using likely	256
Decoding	256
Constructing	259
Likely Comparison Grammar: Examples of all Types	261
Conversions into Likely comparisons	262
Common Pitfalls in Likely Comparisons	263
Forming Comparisons: Tables and Graphs	266
Confounding: A Review	267
Chapter Summary	268
Optional Topics	268
Cases (Deaths) Attributable: Multiple Steps	270

Learning Outcomes

1. Calculate the percentage and counts that are attributable to group membership.
2. Distinguish common-part and distinct-part comparisons
3. Write ordinal and arithmetic comparisons involving named-ratio grammar.
4. Compare ratios in ordinary English using percentage, rate or chance grammar.
5. Compare part-whole ratios using "Likely" grammar.

Review

In chapter 1, we noted that statistics are numbers in context. Statistics can be influenced. The influences were classified into four groups: Confounding, Assembly, Randomness and Error/Bias. "Take CARE" reminds us of these.

In chapter 2, we formed comparisons. Review the arithmetic and the wording.

Table 79 Arithmetic Comparisons

ARITHMETIC COMPARISONS	
Difference	Test is (T-B) more|less than Base
Ratio	Test is (T/B) times as much as Base
% difference	Test is [(T-B)/B]*100% more|less than Base
Times difference	Test is [(T-B)/B] times more|less than Base

Chapter 3 identified three easy ways to mentally control for confounding: the influence of a related factor. They were selection, ratios and standardization.

Ratios are the workhorse in the statistical literacy stable. Chapter 4 introduced *percent* and *percentage*. Chapter 5 introduced *rate* and *chance*. Review these.

Table 80 Named Ratio Grammars

CHAPTER 4 NAMED RATIO GRAMMARS		CHAPTER 5 NAMED RATIO GRAMMARS	
Percent	X% of <Whole> are <Part>	Rate (phrase)	The <Part> rate is n per N <Whole>
	Among <Whole>, X% are <Part>		The rate of <Part> is n per N <Whole>
Percentage	The percentage of W who are P is X%	Chance	The chance of <Process><Result> is X%
	Among W, the percentage who are P is X%		<Process's> chance of <Results> is X%.
	Among W, the percentage of P is X%		The chance that <Process> will <Result> is X%

You need to be very familiar with comparisons and named ratios in this chapter.

Introduction to Comparisons of Ratios

This chapter introduces new arithmetic (the 'attributable to' comparisons) and another named ratio grammar: *likely to*. To repeat: small differences in syntax (grammar) can create big differences in semantics (meaning).

After this chapter, you should understand the difference between these claims:

"216,000 deaths *attributed to* obesity" versus "216,000 deaths *caused by* obesity".

"Among NFL players, blacks *charged with more* crimes than whites" versus "Among NFL players, blacks *less likely to be charged with* crimes than whites."

Chapter 7. Interpreting Confusing Ratios

TABLE OF CONTENTS

Learning Outcomes 271
Review 272
Diabolical Denominators 272
Error: Confusion of the Inverse 275
Medical Tests: Equivocation 277
Social Statistics: Prediction and Explanation 283
Three-term ratios 287
Confounding and Simpson's paradox 288
    Smoking and Lung Cancer 288
    Simpson's paradox for rates and percentages 289
    Cornfield Conditions 290
    Other examples 291
    Controlling for Confounding using Mix Matching 295
The Biggest Confounder 300
    Confounding in the Humanities and Social Sciences 302
Conclusion 303
Optional 304
    Comparing Two Ratios with Related Parts 304
    Comparing Two Ratios with Related Wholes 304
    Distinct and common parts 304
    Unclear wholes 307
    Comparing a Change in Ratios with Different Wholes 308
    Medical Tests: Sensitivity vs. Specificity 311
    Cornfield's Second Condition 312
    Co-Variation involving Percentages 314
    The Scanlan Paradox 315
    Converting Multiple Groups into Two Groups 316

Learning Outcomes

1. Identify and evaluate inverse ratio claims (the confusion of the inverse)
2. Distinguish different kinds of accuracy and error in medical tests.
3. Calculate predictive accuracy given test accuracy and disease prevalence.
4. Distinguish different kinds of statements involving social statistics.
5. Understand the Cornfield conditions: the necessary conditions to nullify or reverse an association.
6. Understand Simpson's paradox
7. Standardize percentages by mix-matching: algebraic and graphical

Review

In chapter 1, we noted that statistics – unlike numbers – can be influenced. In CARE, recall that Error included 'wrong order'. See Figure 198. For ratios, this is known as the confusion of the inverse.

Figure 198: CARE: Classifying Error

CARE: ERROR

Wrong Order	Bias	Lies
Subtract, Divide	Subject	Mistakes
Comparisons	Measurement	Prevarication
Ratios	Sampling	Weasel words

In chapter 2, we studied two group comparisons. We learned that study design (experiments) can protect statistics from various types of confounders.

Figure 199: Control Confounding: Control Of and Control For

CONTROL OF CONFOUNDERS	
Physical Control (Grade = Quality)	
Experiment	Observational Study
A+ Scientific	C Longitudinal
A- Random Assign	D Cross-sectional
B Quasi-Exper	F Anecdotal story

CONTROLLING FOR CONFOUNDERS	
Take into account (mental)	
Can do by hand	*Calculator/Computer*
1 Select/Stratify	4 Linear Regression
2 Form Ratios	5 Logistic Regression
3 Standardize	6 Multivariate Regress

In chapter 3 we studied some of the easier mental ways to control for confounders. Selection and forming ratios are the simplest.

In chapters 4 and 5, we studied ratios of counts using the named ratios. We recognized that the confusion of the inverse is a major problem. In chapter 6 we studied comparisons of ratios.

All of this is background to this chapter. Chapters 3-6 are like calisthenics. This chapter and the next are the payoff! It's time to apply what you've learned.

Diabolical Denominators

Mathematically, a denominator is just another number. School students work problems changing the units of denominator: from inches to centimeters or from pounds to kilograms. But linguistically, the denominator is much more important. It may seem innocent or trivial, but as you will see, it is almost diabolical.

Diabolical denominator: Changing the denominator can change the direction of an association between two ratios. See Table 85:

Table 85: Annual Expenses per Household (HH) and per HH Reporting

Microwave/year	1980	1987	Chg
per household:HH	$14	$20	43%
Percent reporting	2.9%	8%	176%
per HH reporting	$483	$250	-48%
www.bls.gov/opub/mlr/1992/05/art3full.pdf			Table 1

Babysitting / year	1986	1987	Chg
per household:HH	$76	$75	-1%
Percent reporting	6.8%	6.4%	-6%
per HH reporting	$1,118	$1,172	5%
www.bls.gov/opub/mlr/1992/05/art3full.pdf			Table 3

Chapter 8. Randomness

"If there is a 50-50 chance that something can go wrong, then 9 times out of 10 it will." Paul Harvey News, Fall 1979.

Learning Outcomes	317
Review and Background	318
Idea of Statistical Significance	318
Good and Bad of Statistical Significance	319
Influencing Statistical Significance	321
Influenced by Confounding	321
Influenced by Assembly	323
Influenced by Error or Bias	324
Statistical Significance and Unlikely Outcomes	325
50% Chance the Alternate is True	326
Unlikely that the Alternate is True	327
Political Polls	328
Random Samples	330
Margin of Error in Subgroups	332
Survey Margin of Error	332
Statistical Significance Stories	334
Statistical Significance: Applicability	336
Unexpected if Due to Chance: Evolution	337
Common Misunderstandings Involving Chance	338
Chapter Conclusion	339

Learning Outcomes

1. Understand the good and bad of statistical significance

2. See when statistical significance can be influenced – transformed from significance to insignificance or vice versa – by confounding, assembly and bias

3. Calculate confidence intervals given the margin of error for subgroups
   - Calculate the margin of error for subgroups
   - Determine if a difference in subgroup means is statistically significant

4. Distinguish polls from surveys. Polls predict; surveys summarize.

5. Analyze statistical significance impact on clinical trials

6. Identify some common misunderstanding involving chance

Review and Background

Chapter 1 showed how randomness can influence statistics in three areas: extremes, big data and small samples.

1. Extreme outcomes are often just coincidence and can't be replicated.
2. Big data makes it more likely that unlikely outcomes will occur.
3. Small samples make it easier for random variation to influence the results.

Chapter 2 introduced three quantitative ideas involving randomness:

#1: Law of Very Large Numbers. Qualitatively: The unlikely is almost certain given enough tries. Quantitatively: If an event has one chance in N and there are N tries, then one event is expected: the chance of at least one event is more likely than not.

#2: The margin of error is the expected error in random sampling. To form a confidence interval, add and subtract the margin of error to the sample statistic.

#3: A sufficient condition for statistical significance between the means or proportions of two samples is the lack of overlap in their confidence intervals.

Statistically significant describes an outcome that is very unlikely if due just to chance. Statistical significance is evidence for: 1) treating a sample association as real (not spurious) in the population, and 2) treating the difference in a randomly controlled trial as caused by the treatment.

Idea of Statistical Significance

One measure of randomness in the everyday media is the phrase 'statistical significance' or 'statistically significant'. Yes, we don't encounter it very often, but there are times when it is extremely important. Approving new vaccines requires that the results be statistically significant. So this chapter begins with that idea.

Statistical significance is a very abstract idea. It deals with situations that are unlikely. This combination makes it difficult to understand. Let’s think about the idea outside of statistics.

Suppose you are in a relationship. You’ve left a call and three texts with no answer. Normally, you get quick replies. An hour passes, then three hours, then ten. At some point you begin to think. Has something happened? Was there an accident? Is the relationship over? At some point the time delay becomes ‘significant’. It is very unlikely: beyond what is normal, abnormal.

Suppose you are a manager. You’ve learned through experience, that you can’t afford to deal with every problem that crosses your desk. You’ve learned ‘management by exception’: a "policy by which management devotes time to investigating only those situations in which actual results differ significantly from planned results. " This idea was propounded by Frederick Winslow Taylor.

Acknowledgements

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, Leonard Peikoff and Richard Connell), my entrepreneurial guides (Wilbur, Vern and Marshall Schield, George Hock, Signe Schield and John Cerrito) and my statistics teachers (Gerald Kaminski, Peter Holmes and Don Macnaughton).

To my statistical guides: some used statistics as evidence (Charles Murray, Thomas Sowell, Julian Simon, John Lott and Gerald Bracey) and some analyzed statistics as evidence (Freedman et al., Vic Cohn, Hans Zeisel, John Brignell, Jessica Utts, J. H. Abramson, Stanley Lieberson, Robyn Dawes and Howard Wainer).

To Augsburg University for approving two catalog courses: Statistical Literacy (GST 200 in 1997) and Statistical Literacy for Managers (MIS 264 in 2011).

To Bob Hogg who said at JSM 1995: "I'm tired of people talking about problems in the introductory course. I know there are problems. I've written about some of the problems. I want someone to come up with a solution. I want them to write up their solution so I can see it and try it for myself. Then I will know how good it is".
I vowed then that I would write that book. Bob, this book is for you.

To my W. M. Keck project design guides (Magda Paleczny-Zapp and John Knight), my partners (Donald Rubin and Judea Pearl), and my project supporters (Lynn Steen, Peter Holmes, David Kelley, Phil Shively and Allan Rossman).

To the Augsburg project observers who reviewed these materials in class (Linda Schield, Lena Zakharova, Bill Jasperson, Julie Naylor, Boyd Koehler, Peggy Cerrito, Bob Korn, and Cynthia Schield), and to those who reviewed these materials (Larry Copes, Merlin Jetson, Tamra Mason and Robert Geibitz).

To Joel Best, the author of *Damned Lies and Statistics*, for his quintessential contribution—that all reality-based statistics are socially constructed.

To my friend and colleague (Tom Burnham) for 30 years of brutally clear thinking and to my colleague (Marc Isaacson) for 15 years of helpful ideas and feedback. To Alison Oliver for encouraging me to submit my book idea to Wiley for publication. To Carl Lee for nominating me as an ASA Fellow[325] and to Danny Kaplan, Jeff Witmer, Chris Wild, Herb Weisberg and John Bailar for supporting my application. To Erik Erhardt for choosing to implement this course at the Univ. of New Mexico.

A very special thanks to the W. M. Keck Foundation – and to Mercedes Talley in particular – for their 2001 grant "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 possibility. A most special thanks to my wife, Cynthia, who encouraged me to get this book done.

[325] Schield named ASA Fellow: www.StatLit.org/pdf/2018-Schield-ASA-Fellow.pdf

Table of Figures

Figure 1: Important things to know about statistics 10
Figure 2: An Argument is Like a House 12
Figure 3: Minutes to Defibrillation vs. Heart Attack Survival 13
Figure 4: Imported Lemons versus Highway Fatalities 13
Figure 5: CARE identifies the four kinds of factors that influence a statistic....14
Figure 6: Statistics as Evidence in an Argument: Take CARE 15
Figure 7: Distinguishing critical thinking from statistical literacy 56
Figure 8: Association (Statistical): Comparison vs. Co-Variation 16
Figure 9: Association Diagrams 17
Figure 10: Association Shown Graphically....17
Figure 11: Arrows Indicating Causation 19
Figure 12: A-B-C Grammar....19
Figure 13: Association is typically evidence of causation (somewhere)....20
Figure 14: Disparities are not [necessarily] Discrimination 21
Figure 15: Prevalence of Suicide by German Provinces in the 1500s....22
Figure 16: Argument as a House 25
Figure 17: House Graphs for Three Arguments....25
Figure 18: Statistics as Evidence in an Argument (House) 26
Figure 19: Triangle Diagrams Involving Confounding 28
Figure 20: Take CARE: Confounders by Type of Association 28
Figure 21: Common-Cause Triangle: Ice Cream & Burglaries 29
Figure 22: Single-Cause Triangles: Sickle-cell anemia & Better economy....29
Figure 23: Triangle Diagram of a Confounder: Golf Courses & Divorces 30
Figure 24: Down syndrome: Increase by Birth Order and Mom's Age....31
Figure 25: Story Diagrams for Father and his Kids 31
Figure 26: CARE: Classification of Assembly: A top-level overview....32
Figure 27: Global Temperatures since 1940 (left) since 1820 (right) 36
Figure 28: CARE: Classify Randomness; Sports Illustrated Jinx 37
Figure 29: After 1st try: Best did worse; Worst did better....38
Figure 30: Random Characters Generating Words 39
Figure 31: Galton Board (Bean Machine)....41
Figure 32: CARE: Classifying Error 42
Figure 33: Class Scores versus Class Size (Random data) 44
Figure 34: Bias: Survivor bias (left); Bias vs imprecision (right) 45
Figure 35: Error in tests: false positive versus false negative....47
Figure 36: Take CARE Influences on a Statistic 50
Figure 37: Chapter 1 key words, phrases and ideas 50
Figure 38: Inferences from Observed to Unobserved....52
Figure 39: Classifying third factors to an Association....55
Figure 40: Triangle Diagram of a Mechanism: Lighting Causes Thunder....55
Figure 41: Runs from random flips of a fair coin 56
Figure 42: Randomly Scattered Grains of Rice 57
Figure 43: Randomly-generated birthday date matches 58
Figure 44: Six Ordered Comparisons 61

Figure 45: Template: Four Kinds of Arithmetic Two-Group Comparisons....... 63
Figure 46: Template: Four Grammatical Forms of Co-Variation....................... 68
Figure 47: Confounder Solutions: An Overview 69
Figure 48: Studies: Controlled vs Uncontrolled.. 70
Figure 49: All Studies: Common & Technical Names 72
Figure 50: Random Assignment 73
Figure 51: Random Assignment Nullifies Prior Confounding 74
Figure 52: Salk vaccine clinical trial 74
Figure 53: Quasi-Experiments: Types Of... 75
Figure 54: Influence of Advertising on Consumer Response 76
Figure 55: Salk vaccine: Random assignment vs. quasi-experiment................. 76
Figure 56: Cigarette consumption and lung cancer deaths 77
Figure 57: Health Benefits of Smoking Cessation 77
Figure 58: Cross-sectional studies involving comparison and covariation 78
Figure 59: Golf scores versus Age.. 78
Figure 60: Studies: Controlled vs Uncontrolled... 79
Figure 61: Experiments vs. Observational Studies: Ranking............................. 80
Figure 62: Two Types of Control 82
Figure 63: Assembly/Assumption Solutions (An overview) 85
Figure 64: Antarctic Temperature Data 88
Figure 65: CARE: Classification of Randomness.. 91
Figure 66: Business Performance: Before and After a Given Year................... 91
Figure 67: Meaningful words from randomly-generated letters 93
Figure 68: Test Scores vs. Class Size: Above Average and All Scores 96
Figure 69: Take CARE: Error... 98
Figure 70: Summary of Arithmetic Two-Group Comparison Grammar.......... 104
Figure 71: Chapter 2 Keywords, Phrases and Ideas... 104
Figure 72: Study Design: Control Of Confounders.. 104
Figure 73: Scientific Experiment: Before/After Brief Run.............................. 106
Figure 74: Repeated Measures of Pulse Rates (Longitudinal)......................... 110
Figure 75: Longest Run of Heads in Flipping Fair Coins 111
Figure 76: Structures formed by grains of rice dropped randomly.................. 112
Figure 77: Birthday problem: Random birth dates for 28 people at a table. 112
Figure 78: Column Charts of Discrete Categories or Values........................... 115
Figure 79: Continuous Data: Histogram and Column Chart............................ 116
Figure 80: Shapes of Distributions 117
Figure 81: Distribution of Student Grades and House Prices 118
Figure 82: Cumulative Distribution of Houses by Price 119
Figure 83: Percentiles on ACT and SAT Tests 120
Figure 84 Distribution of House Prices: Mean... 122
Figure 85 Distribution of House Prices: Median.. 122
Figure 86 Distribution of House Prices: Mode... 123
Figure 87 Centers for Symmetric Unimodal Distributions 124
Figure 88 Centers for Right-Skewed Distributions.. 124
Figure 89 2000 U.S. Distribution of Deaths by Age and Sex 125

Figure 90 US mass shootings and Mean Verbal SAT scores 127
Figure 91 Heart Attack Prevalence vs. Cholesterol Level 130
Figure 92 Confounders: Control Of and Control For..132
Figure 93 Take Into Account Confounder Influence 132
Figure 94 U.S. Income Distribution: Before/After Adjust...............................133
Figure 95 State NAEP Scores Confounded by Internet Access 137
Figure 96 Weighted-Average Graph for Silverware (1/3)................................142
Figure 97 Weighted-Average Graph for Silverware (2/3)................................143
Figure 98 Weighted-Average Graph of Silverware (3/3) 144
Figure 99 Triangle for Average Cost of Silverware..145
Figure 100 LA vs. WV NAEP Scores (Raw)..146
Figure 101 LA vs. WV NAEP Scores (Standardized) 147
Figure 102 Standardizing Race-Based Incomes: Step 1 149
Figure 103 Standardizing Race-Based Incomes: Step 2 149
Figure 104 Standardizing Race-Based Incomes: Step 3 150
Figure 105 Summarizing the Math scores example..153
Figure 106 Summarizing the black-white family income gap.........................153
Figure 107 Major League Wins vs. Total Payroll...154
Figure 108 Distribution of IQs (left) and Boy-Girl Ratios by IQ (right) 156
Figure 109 Chapter 3: Keywords, Phrases and Ideas..157
Figure 110 Take CARE and Components of Error 161
Figure 111 Controlling Confounders: Control Of and Control For..................161
Figure 112 Preposition-Based Ratios and Named Ratios 162
Figure 113: Controlling for relevant confounders...163
Figure 114 Pie Charts: OK (left) and bad (right) 164
Figure 115 Pie Charts: Smokers (left) and Students (right)..............................164
Figure 116 Percent grammar questions 165
Figure 117 Template for Percent Grammar Statements....................................166
Figure 118 Part-Whole Grammars: General Decoding Rules 168
Figure 119 Percent Grammar Statements: Specific Rules 168
Figure 120 Answer Part-Whole Questions using Count Table Data 168
Figure 121 Create Part-Whole Ratio Table from a Count Table 168
Figure 122 Describe a specific Percentage in a 100% Table............................170
Figure 123 Margin Values: Sums or Averages 170
Figure 124 Part-Whole Pie Chart: Age of Drivers in Accidents 174
Figure 125 Template for Percentage Descriptions..177
Figure 126 Using a Pie Chart to Represent a Part-Whole Ratio.......................178
Figure 127 Distribution of Religions by Race (Of Races by Religion).............182
Figure 128 Margin Value Rule 184
Figure 129 Requirements for One-Way Half Tables in Columns 185
Figure 130 Missing Margin Rule...188
Figure 131 Measuring Happiness 190
Figure 132 Percent and Percentage Grammar Summary 191
Figure 133 Chapter 4 Keywords, Phrases and Ideas...192
Figure 134 Controlling for Confounders (Repeat)..192

Figure 135: Controlling confounding: Control of vs control for 198
Figure 136 Preposition-Based Ratios and Named Ratios 198
Figure 137: Template for Rate Descriptions: Phrase-Based 203
Figure 138: Rate Grammar Decoding Rules 204
Figure 139: Pie Diagram: Rate of Accidental Deaths 204
Figure 140: Template for Clause-Based Rate Descriptions 207
Figure 141: Percentage and Rate Grammar in Titles and Comparisons 207
Figure 142: Template for Chance Descriptions: Clause-Based 209
Figure 143: Template for Chance Descriptions: Phrase-Based 209
Figure 144: Wheel of Inference (Repeat) 210
Figure 145: Rules for Decoding Descriptions of Ratios 212
Figure 146: Chapter 5 Keywords, Phrases and Ideas 213
Figure 147: FDA Drug Approval Success Rates 226
Figure 148: US Death Rate 1960-2017: Crude and Age-Adjusted 228
Figure 149: US crude birth rate: 1800-2020 228
Figure 150: Covid19 Infections, Cases and Infection-Case Ratio 229
Figure 151: US Covid19 Daily Deaths-per Case and US TSA Passengers 229
Figure 152: CPI: 2020-2021 229
Figure 153: Interest Rates on US 10 Year Treasury Notes since 1964 230
Figure 154: Interest Rates on US 10 Year Treasury Notes since 2019 231
Figure 155: US Total Mortgage Debt since 1950 231
Figure 156: US Autopsy Rates per 100 Deaths by Cause 232
Figure 157: Distribution of Black Families by Marital Status since 1950 232
Figure 158: Delinquency Rates: Auto Loans and Student Loans since 2003... 233
Figure 159: Consumer debt components since 2003 233
Figure 160: GDP per capita since 1270 and US Federal Debt thru 2020 233
Figure 161 High School Graduation Rates by Gender 234
Figure 162 Distribution of Covid Hospitalizations by Age (2021) 234
Figure 163 Distribution by Vaccination Status (2021) 234
Figure 164 Classification of Households 235
Figure 165 2019 Median Income: Household vs. Family 235
Figure 166 Distribution of Households by Income. PDF & CDF 236
Figure 167 Cumulative Distribution of Households and Total Income 236
Figure 168 Male Testosterone versus Age 237
Figure 169: Crime rates by city size 237
Figure 170: State's Covid Vaccination rates by race 238
Figure 171: Percentage Attributable to an Exposure 242
Figure 172: Common-part comparison of Part-Whole ratios 246
Figure 173: Distinct-parts comparison of Part-Whole Ratios 246
Figure 174: Two Pie Diagram 246
Figure 175: Pie Compare for Example 1: Common Part 248
Figure 176: Pie Compare for Example 2: Distinct-Part Comparison 248
Figure 177: Compare two percentages as a percentage difference 249
Figure 178: Compare two percentages as a difference: larger as the base. 250
Figure 179: Compare two rates using a ratio comparison 250

Figure 180: Compare two probabilities using a ratio comparison 250
Figure 181: Ratio Comparison of Named Ratios: Percentage grammar 251
Figure 182: Percent difference compare of Ratios; Percentage grammar 252
Figure 183: Percent difference comparison of Named Ratios: All 253
Figure 184: Common-Part Ratio Compare Template: Percentage Grammar....253
Figure 185: Distinct-Part Ratio Compare Template: Percentage grammar 253
Figure 186: Creating Short Form Comparisons from Long Forms 254
Figure 187: Common Part Comparison using Two Pie Charts 254
Figure 188: Two Pie Comparisons 255
Figure 189: Rules for Likely Comparisons of Ratios 257
Figure 190: Common Part Comparison using Two Pie Charts 257
Figure 191: Distinct Part Comparison using a single Pie Chart 257
Figure 192: Likely comparison template: common-part compare 259
Figure 193: Likely comparison template: distinct-parts compare 260
Figure 194: MS Word: as likely … than… Suggested corrections 264
Figure 195: Chapter 6 Keywords, Phrases and Ideas 268
Figure 196: Two-Group Comparisons, Named Ratios and Comparing Ratios.268
Figure 197: Deaths Attributable to an Exposure 270
Figure 198: CARE: Classifying Error 272
Figure 199: Control Confounding: Control Of and Control For 272
Figure 200: Covid Data (Michigan vs. Rhode Island) 273
Figure 201: Crimes by Level for Gender and Race 274
Figure 202: Crimes by Level (include Committed) for Gender and Race 274
Figure 203: Leading Causes of Death in 2001 287
Figure 204: Ronald Fisher and Jerome Cornfield 288
Figure 205 Greatest Contributions of Statistics to Human Knowledge 289
Figure 206: Triangle Diagram: Hospital Death Rates and Patient Condition ...289
Figure 207: Death Rates by Hospital and by Patient Condition 291
Figure 208 Death Sentence Rates by Race of Murderer and Victim 292
Figure 209 Magazine Subscription Renewal Rates 294
Figure 210 Basketball Completion Rates by Team and Type of Shot 295
Figure 211: Mix-Matching Graphically: Hospital Death Rates (data) 297
Figure 212: Mix-Matching Graphically: Hospital Death Rates (Crude) 297
Figure 213: Mix-Matching Graphically: Hospital Death Rates (Adjusted) 298
Figure 214: Comparative Binary Separations 301
Figure 215: Chapter 7 Keywords, Phrases and Ideas 303
Figure 216: Compare Suicides for Widows vs. Widowers 306
Figure 217: Prevalence by Hospital: Death vs. Poor-Health 313
Figure 218 Distributions by Month: Renewals vs. Directs 313
Figure 219 Prevalence by Team: Completions vs. Two Pointers 314
Figure 220: Down syndrome: Child's Birth Order and Mom's Age 316
Figure 221: Statistical Significance: Crude Statistics 321
Figure 222: Statistical Significance: Crude and Adjusted Statistics 323
Figure 223: Statistical Significance: Different Definitions 323
Figure 224: Statistical Significance: Biased and Unbiased 324

Figure 225: Random Sampling: Simple and Stratified 330
Figure 226: Random Sampling: Cluster and Systematic 331
Figure 227: Random Sampling: Classification and Evaluation 331
Figure 228: Chapter 8 Keywords, Phrases and Ideas 339

Table of Tables

Table 1: U.S. persons living with AIDS 35
Table 2: U.S. presidential elections where taller candidate won 51
Table 3: U.S. Accidental Deaths by Type of Accident (US CDC) 61
Table 4: Average Weight by Smoking Status and Gender 84
Table 5: US Professional Degrees; Households by Size 115
Table 6 Distribution of Heights for U.S. Twenty-year olds 120
Table 7: Average U.S. Family Income by Quintile and Top 5% in 2003 121
Table 8: 1999 U.S. Family Incomes by Family Type 125
Table 9: Survival from Constant Rate of Decay (Exponential) 125
Table 10: Family Incomes by Highest Education: Median and Mean 128
Table 11: Family Incomes by Wage Earners: Median and Mean 129
Table 12: State Prison Operating Expenses: California vs. New York 134
Table 13: State Prison Operating Expenses: Maryland vs. Kansas 135
Table 14: State Prison Operating Expenses: Minnesota vs. Maine 135
Table 15: State Prison Operating Expenses: Minnesota vs. Iowa 135
Table 16: NAEP 2000 8th Grade Math Scores: VA vs. TX 136
Table 17 SAT Verbal Scores by Race: 2002 vs. 1982 137
Table 18: NAEP Scores 4th Grade Math (Crude): LA vs. WV 139
Table 19: NAEP Scores 4th Grade Math (Adjusted): LA vs. WV 140
Table 20: Mean Family Incomes (Crude) by Race & Marital Status 140
Table 21: Mean Family Incomes (Adjusted) by Race & Marital Status 140
Table 22: SAT Verbal Scores by Race: Crude 141
Table 23: SAT Verbal Scores: Standardize 2002 scores on 1982 weights 141
Table 24: Silverware Average Costs: Crude and Adjusted 145
Table 25: NAEP Scores 4th Grade Math: LA vs. WV 145
Table 26: Mean Family Incomes by Race & Marital Status 148
Table 27: Family Income by Race: Crude vs. Adjusted 151
Table 28: Toy Table of College Students 165
Table 29: Toy Table of Students: Row Percentages 168
Table 30: Sample Table of Row Percentages 169
Table 31: Sample Table of Column Percentages 169
Table 32: Sample Table of Total Percentages 169
Table 33: World Population by Religion and Continent (1996) 171
Table 34 U.S. Distribution of First Marriages by Woman's Age (%) 172
Table 35 U.S. Women: Distribution by Contraceptive Status 173
Table 36 U.S. Age of Licensed Drivers and Drivers in Accidents 174
Table 37: Sample Table of Column Percentages 178
Table 38: Sample Table of Row Percentages 179
Table 39: U.S. Teacher Status: 2004-05 vs. 1988-89 179
Table 40 U.S. Lifetime Mobility by Religious Preference 180
Table 41 Sample One-Way Half Table 183
Table 42 Sample: Percentage Uninsured with Helpful Titles 183
Table 43 Percentage Uninsured with Unhelpful Titles 184
Table 44 Cigarette Smoking 185

Table 45 Moral Acceptability of Activities by Age Group 185
Table 46 Full Three-variable Table: Row-Percentages 186
Table 47 Smokers Two-Way Half Table (Helpful title) 186
Table 48 Smokers Two-Way Half Table (Unhelpful title) 187
Table 49 Percentage overweight or obese 187
Table 50 U.S. 2004 School Status Dropouts of 18-24 year olds 188
Table 51 Percentage of Marijuana or Hashish Users by Age 189
Table 52 U.S. Percentage of Smokers by Age 189
Table 53 Abortion Percentages by Party 190
Table 54: U.S. Current (last month) Drug Use Percentages 190
Table 55 U.S. 1990 Accidental Deaths by Race and Sex (Frequency) 201
Table 56 U.S. 1990 Population by Race and Sex (Count) 201
Table 57 U.S. 1990 Accidental Death Rates by Race and Sex (Incidence) 202
Table 58 Auto Traffic Death Rates in the U.S. 205
Table 59 Distribution of Households by Family Structure and Race 215
Table 60 Historical data on divorced women who remarry 216
Table 61 Distribution of US Income and Federal Tax: 1980-2018 216
Table 62 Factors Influencing Income Distribution 217
Table 63 Income Mobility 217
Table 64 Total Interest Paid as a percentage of the Loan 218
Table 65 1860 US Census: Slaves by geographical region 219
Table 66 1860 US Census: Distribution by Slaves per Household 219
Table 67 Distribution of Slave-Owning Households and Slaves 220
Table 68 Percentage of US children who live with a single parent by race 220
Table 69 NYC Race/Ethnicity Distribution: Foster Care vs. Population 221
Table 70 Number of Murder Victims by Race, Sex and Age 221
Table 71 Law Officer Killed by Race and Sex of Offender 221
Table 72 2019 Percentage of victimizations reported to police 222
Table 73 Violent crime rates: reported and unreported 222
Table 74 Violent incidents by sex and race of victim and offender 223
Table 75 Educational Attainment of those Formerly Incarcerated 223
Table 76 People killed by the police: counts and percentages 224
Table 77 Family Net Worth and Income: Mean and Median 225
Table 78 Family Income by Characteristic: mean and median 225
Table 79 Arithmetic Comparisons 240
Table 80 Named Ratio Grammars 240
Table 81: Students by Sex and Smoking Status 251
Table 82 Percentage who are Runners by Sex and Smoker 252
Table 83: U.S. Unemployment Rates by Highest Grade, Sex & Race 252
Table 84: IQ and Prison: Toy table of Counts 269
Table 85: Annual Expenses per Household (HH) and per HH Reporting 272
Table 86: Covid19 Death Rates: Michigan vs. Rhode Island 273
Table 87: DNA Match: Confirmation vs. Prediction 277
Table 88: Outcomes of a Medical Test 278
Table 89: Quality of a Medical Test in Confirming (column) 278

Table 90 Quality of a Medical Test in Predicting (row) 279
Table 91 Medical Test, Blank Form for Counts 280
Table 92 Medical Test to Detect HIV, 95% Accurate 0.1% Prevalence 280
Table 93 Medical Test to Detect HIV, 95% Accurate, 1% Prevalence 281
Table 94 Medical Test to Detect HIV, 99% Accurate, 1% Prevalence 281
Table 95 Medical Test to Detect HIV, 95% Accurate, 50% Prevalence 282
Table 96 Gave Birth in Last Year by Gender among Young Adults 283
Table 97 Toy Table: Police Calls, 90% Accuracy; 10% Prevalence 284
Table 98 Toy Table: Police Calls, 90% Accuracy; 50% vs. 90% Prevalence...285
Table 99 Low IQ (≤90) and Prison 285
Table 100: Suicide-Bulimia Connection (Toy data) 286
Table 101: Male juveniles: Abuse Crime Connection (Toy data) 286
Table 102: Mortality by Hospital 289
Table 103: Mortality by Hospital and Patient Condition 290
Table 104: Death Sentence Rates and Counts by Race of Murderer 292
Table 105: Death Sentence Rates and Counts by Race of Victim 292
Table 106: Death Sentence Data by Race of Murderer and Victim 292
Table 107 Renewal Rate by Division and Month 293
Table 108 Renewal Rate by Division 293
Table 109 Basketball Field Goals by Team & Shot 295
Table 110 Basketball Field Goals by Shot 295
Table 111 Mix-Matching Arithmetically: Hospital Patient Death Rates 296
Table 112 Mix-Matching: Murderers' Death Sentences 298
Table 113 Mix-Matching: Magazine Renewal Rates 299
Table 114 Mix-Matching: Basketball Shots Made 299
Table 115 Percentage who smoke by sex, weight, runner and height 301
Table 116 1990 U.S. Accidental Death Rates by Race and Gender 301
Table 117 U.S. Poverty Rates by Race and Age 301
Table 118: Single Parents by Race and Ethnicity 306
Table 119 Registered Adults vs Registered Voters (toy data) 308
Table 120 Infant Deaths per 1,000 Births by Cause: 1920 vs. 1960 309
Table 121 Greenhouse Gas Concentrations 310
Table 122 Medical Tests: Pap Smears 312
Table 123: Mortality of Hospital Patients by Condition 312
Table 124 Subscriptions Due by Division and Month 313
Table 125: Infant Mortality Rates: 2000 vs. 1990 315
Table 126 Percentage of Births which are Low Weight (toy data) 322
Table 127 Hypothesis test as Medical Test: 50% Prevalence 326
Table 128 Hypothesis test as Medical Test: 5% Prevalence 327
Table 129: Hypothetical Presidential Poll Data; Distribution of Voters 333

Table of Stories

Story 1: Fruits, Veggies Cut Risk of Breast cancer... 8
Story 2: Two Hunters and a Bear.. 25
Story 3: Father's kids out of control.. 27
Story 4: Child Abuse in Minnesota: 35
Story 5: Every 14 seconds, AIDS turns a Child into an Orphan......................... 36
Story 6: Similarities between Presidents Lincoln and Kennedy......................... 40
Story 7: IQs of U.S. Presidents 45
Story 8: Self-selection bias: Surveys 46
Story 9: Risk of Dying (generic).. 47
Story 10: Unfair boss could shorten your life: study 48
Story 11 Galileo Tower of Pisa.. 73
Story 12 Harvey Circulation of Blood.. 73
Story 13 Random Assignment: Lady Tasting Tea: 74
Story 14: Salk vaccine reduces Polio Risk 76
Story 15: Study: Placebos Make People Feel Better... 98
Story 16: Placebo Usage Criticized 100
Story 17: Do Magnets Reduce Pain? 101
Story 18: Teens who eat breakfast weigh less... 102
Story 19 Random Assignment: Tinctures as Cures... 107
Story 20 Random Assignment: Cure for Tuberculosis 108
Story 21: Quasi-Experiment: Lemons and Scurvy.. 108
Story 22: Quasi-Experiment: Cowpox and Small Pox....................................... 108
Story 23: Quasi-Experiment: Prontosil Reduces Mother's Risk of Death 108
Story 24: Does Smoking Increase the Risk of Lung Cancer?........................... 110
Story 25: Percentiles Involving Admission to Law School 121
Story 26: U.S. Distribution of Incomes 133
Story 27: Do You Worry More About Your Health?....................................... 265
Story 28: What to do when you test positive for HIV...................................... 279
Story 29: Suicides: More Widows than Widowers 305
Story 30: Increase in Health Costs.. 307
Story 31 Each Daily Soda Increases Obesity Risk 60% 314
Story 32: Diabetes drug effective 334
Story 33. Duct tape no magical cure for warts, study finds (Generic).............. 335
Story 34. A Double-Blind Test of Astrology 335

Table of Equations

Eq. 1 N times as much as = (N-1)*100% more than..65
Eq. 2 N times as much as = (N-1) times more than 65
Eq. 3 Confidence Interval: Sample Statistic ± Margin of Error 97
Eq. 4 Mean = Sum of value / # of subjects 122
Eq. 5 Weighted Ave: (Wt1*Ave1 + Wt2*Ave2 + …) / Sum(Weights)............138
Eq. 6 Weighted Ave: (Ave1*Associated fraction + Ave2 * Associated fraction/1 138
Eq. 7 Percentage difference explained: 100% (Before – After)/Before............150
Eq. 8 Part-whole percentage = (#Part / #Whole) x 100%.................................166
Eq. 9 Percentage attributed to exposure=100%*(ExpRate-CtrlRate)/ExpRate 242
Eq. 10 Cases attributed to exposure = % Attributed * # cases exposed 244
Eq. 11 ME(subgroup) = ME(group)*SQRT(1/Fraction) 332

Glossary

100% Sum Rule	100% Sum Rule: If a margin value is a 100% sum, then the group is whole and each component is a part. P. 171
Accuracy: Confirmation	Confirmation accuracy in a medical test is the percentage of diseased cases that test positive; the percentage of disease-free that test negative. P. 278
Accuracy: Prediction	Prediction accuracy in a medical test is the percentage of positive outcomes that involve the disease (fraction of negatives that are disease free). P. 278
Alternate explanation	An alternate explanation for an association can be provided by a confounder but never by a mechanism. P. 55
Ambiguous	Ambiguous means vague on essentials. P. 19
Anecdotal evidence	Anecdotal evidence involves a study or story about an individual or a small group based on limited experience. P. 72
Apples+oranges comparison	See crude comparison or mixed-fruit comparison
Arbitrary evidence	Arbitrary evidence is evidence that is purely hypothetical or is weakly related to the argument. P. 26
Argument: sound	A sound argument is one that is valid and has true premises. P. 105
Argument strong	A strong argument is one where the premises give strong support for the conclusion.105
Argument: valid	A valid argument is one where the conclusion must true given that the premises are true. P. 105.
Argument: weak	A weak argument is one where the premises give weak support for the conclusion. P. 105
Arithmetic comparison	Arithmetic comparisons compare two values by showing the direction and size of the comparison. Arithmetic two-group comparisons are of four kinds: difference, times ratio, percent difference and times difference. P. 62
Assembly	Assembly involves all the choices that influence a statistic and are not covered elsewhere under CARE. P. 32
Assembly solutions	Assembly solutions involve maturity and critical thinking: close-reading, hypothetical thinking, estimating the impact and asking questions. P. 85
Association	Association (statistical) is a quantitative connection between groups or characteristics. P. 16

Association: crude	A crude association is one that fails to take into account plausible confounders. Also known as a mixed-fruit association or an apples and oranges association. P. 87
Association: flimsy	A flimsy association is one that is readily influenced. P. 105
Association: solid	A solid association is one that is resistant to being influenced (not readily influenced). P. 105
Association words	Association words assert association explicitly or describe associations involving fixed conditions or unrepeatable events. P. 19 See also Causation words and Between words.
Availability bias	Availability bias: A judgement about an event or statistic that is influenced by how readily one recalls examples. P. 99
Average	Average: See Mean.
Bar chart	A bar chart is a chart that uses bar lengths or heights for the amounts. The bars can be vertical (column chart) or horizontal (row charts). The bars can be separated or touching. P. 116
Base part	A base-part is a base in a comparison and a part in a part-whole ratio. P. 246
Base whole	A base-whole is a base in a comparison and a whole in a part-whole ratio. P. 246
Bayes comparison	A Bayes comparison is a common-part ratio comparison in a group versus the entire population. It allows an exchange of the part with the test whole with no change in the numerical strength of the ratio comparison. P. 269
Between words	Between words are words whose meanings are 'between' association and causation. They describe an association but suggest causation. P. 20
Bias	Bias is systematic error: a "systematic deviation of results or inferences from the truth." See Respondent bias, sampling bias and measurement bias. P. 43
Big data	Big data (statistically) is any data set that is so big that all the associations are statistically significant. P. 92
Causation words	Causation words assert causation, state sufficiency or state a contra-factual.. P. 20
Cause	A cause is an event or condition whose level or presence makes (or can make) a difference in something else. P. 18
Cases attributable	Cases attributable to exposure are those deaths in the exposed group that are attributable to being in the exposure group. P. 243

Centers rule	Three centers rule: The three measures of center have a natural order in most skewed distributions. It is alphabetic in English: mean, median and mode if skewed left; mode, median and mean if skewed right. P. 124
Chance	Chance is a possibility of something happening. P. 37
Chance grammar	Chance grammar describes part-whole ratios using the keywords *chance, risk, odds, probability and likelihood*. P. 208
Cherry picking	Selection bias: Selecting just those subjects or data that support your claim.
Clinical trial	Randomized Controlled Trial (RCT) is a controlled experiment (trial) involving random assignment. *Clinical* trial is *the* most common equivalent. P. 73
Close reading	Close reading involves seeing how the presence, choice or absence of a single word or phrase could have a big impact on the statistic or statistical association. Small changes in syntax can create big changes in semantics. P. 85
Coincidence	Coincidence is a noteworthy connection between unlikely events with no obvious connection. P. 37
Columns, rows	Columns run vertically like columns in a building. Rows run horizontally like rows of seats in a theater. P. 165
Common-part comparison	A common-part comparison compares two ratios having a common part of two different wholes. P. 245
Common part	A common-part exists when two part-whole ratios share the same part. P. 245
Comparison bias	Comparison bias occurs when prior differences between the control group and the treatment/exposure group influence the comparison after a treatment/exposure. P. 99
Comparison: mixed-fruit	Mixed fruit comparison: a crude comparison of groups with different mixtures of a related variable. Also known as an 'apples and oranges' comparison. P. 136
Comparison: arithmetic	Arithmetic comparisons compare two values by showing the direction and size of the comparison. P. 62
Comparison: ordered	Ordered comparisons compare two values without mentioning their size or the size of the comparison. P. 60
Comparison: raw	Raw comparisons just present the values to be compared. P. 60
Completion bias	Completion bias: an outcome that is influenced differently by those who do and don't complete. P. 99

Component	Components are cells within a group: a row, a column, or a range. P. 171
Confidence interval	A confidence interval is an interval around the sample statistic (SS) such that 95% of these intervals contain the associated population statistic. P. 97
Confirmation bias	Confirmation bias: selecting just the data that agree with one's thinking, or repeating/confirming an idea so much that it is accepted as true. P. 99
Confounded	Confounded means confused; *confounding* means confusing. P. 27
Confounder	A confounder is a related variable that was not included in generating the association that provides an alternate explanation for an association. Technically a confounder is a third factor that causes the result in an association, and that is related to – but not caused by – the predictor. P. 28
Confusion of the inverse	Confusion of the inverse is a grammatical error where two variables in a ratio are mistakenly exchanged without changing the number. P. 275. The confusion of the inverse exchanges part with whole in a part-whole ratio. P. 170
Context	Context: the relevant circumstances: the reality. P. 11
Control of; Control for	'Control of' is physical; to assign subjects to treatment or control groups, to set the values of relevant factors. P. 82 'Control for' is mental: to take into account or adjust for. P. 82
Control group	The control group is the group not treated or exposed. P. 70
Controlled study	A controlled study is any study that involves at least two groups where one is treated or exposed; the other is not. P. 70
Convenience sample	A convenience sample (a grab sample) is a sample "selected by easily-employed non-random methods." P. 95
Cornfield condition #1	Necessary Condition #1: Simpson's paradox occurs only if the association is bigger for the confounder-outcome than for the predictor-outcome. E.g., the death rate comparison is bigger by patient condition than by hospital. P. 290
Cornfield condition #2	Necessary condition #2: Simpson's Paradox occurs only if the predictor-confounder association is bigger than the original predictor-outcome association. E.g., the predictor- confounder (patients in poor condition) association is stronger than the hospital-death rate association. P. 312
Cross-sectional study	A cross-sectional study is an observational study that involves a single moment in time (unemployment rate) or over a time interval (death rate). P. 72

Crude comparison	Crude comparison: a comparison that conflates (fails to take into account) important related factors. Also known as a mixed-fruit comparison or an apples and oranges comparison. P. 30
Dashed lines	Dashed lines indicate an association. P. 28
Data: count	Count data counts how many things have a discrete quality: *categorical* (male vs. female), *ordinal* (disagree, neutral, agree) or *quantitative* (number of cars).P. 115
Data: measure	Measurement data measures quantities: characteristics that can have any value within some range. P. 115
Deductive reasoning	Deductive reasoning: reasoning where the conclusion must be true if the premises are true and the argument is valid. Often called formal reasoning. E.g., All men are mortal. Aristotle is a man. Therefore, Aristotle is mortal. P. 24
Determiners	Determiners are conditions that determine or delimit the size of the whole or part. P. 167
Diabolical denominator	Diabolical denominator: Changing the denominator can change the direction of an association between two ratios. P. 272
Discriminate	Discriminate (verb) is ambiguous: to discern a difference (good) or to judge with prejudice (bad). P. 21
Discrimination	Discrimination typically means to judge with prejudice. P. 21
Disparate impact	Disparate impact: an unintended disparity between groups resulting from group-neutral treatments. P. 21
Disparity	Disparity**:** a difference (lack of equality or parity) between groups. P. 21
Distinct-part comparison	A distinct-parts comparison compares two ratios involving two different parts of a common whole. P. 245
Distinct parts	Distinct parts exist when two part-whole ratios share the same whole. P. 245
Doing	Doing is when the researcher manipulates or nature intervenes while controlling the subject and the environment. P. 18
Double-blind study	A double-blind study blinds the researcher as well as the subject from knowing which group a subject is in, thereby eliminating bias for both. P. 100
Double-ratio comparisons	Double ratio comparisons are ratio comparisons of ratios. P. 248
Double-who ambiguity	This ambiguity occurs in percentage grammar when two relative pronouns follow *percentage*. P.181

Effect Size	Effect size is the size of a two-group comparison (typically a times ratio) or the size of the slope in covariation. P. 79
Effectiveness	Effectiveness in preventing an outcome is one minus (the rate in the treatment group divided by the rate in the control group). P. 268
Error	Errors are systematic deviations from what is real or true. P. 42
Evidence	Evidence involves less-disputable claims that support the point of the argument. P. 51
Experiment	An experiment is a study involving a doing: subjects are given or assigned a treatment by a researcher or by nature. P. 18
Explanatory power	The more important variable is the variable that has the higher explanatory power: the stronger association with the variable of interest. P. 300
Explanatory power rule	The explanatory power rule: the more important of two binary predictors is typically the one having the greater effect size: the bigger difference, times ratio or percentage difference in the variable of interest. P. 300
Exposure group	The exposure group is the group in an observational study that is exposed. P. 70
Factor	A factor or variable is a property or characteristic of something. P. 16
Family	A family (in housing) is a household whose members are related by blood or law. P. 128
Frequency (rate)	Frequency or velocity (rate): events per unit time (heart rate or speed. P. 199
Frequency distribution	A frequency distribution gives the count of subjects in each category or in a range of values. P. 115
Generalization	Generalization infers a property of a group (population) based on a related property in sample from that group. P. 52
Group in a table	A group is a row, column or table whose subjects make up the subjects in a margin value. P. 171
Half table	A half table is half of a full 100% table with enough information to complete the table. P. 183. A multiple one-way half table consists of a series of one-way half tables placed side-by-side as columns or rows. P. 183
Halo effect	The halo effect is when the researcher's optimism influences the data to support that optimism. P. 99

Hawthorne effect	The Hawthorne effect is a systematic change in response when the subjects know they are the subject of attention. P. 98
Histogram	A histogram is a bar chart where a bar spans an interval, so the bars can touch. P. 116
hypothesis, research	Research hypothesis: a claim involving a difference or change that the researcher hopes will be true in the larger population.. P. 325
Hypothetical thinking (assembly)	Hypothetical thinking (assembly) is thinking about different ways that a statistic could have been created, compared, adjusted or presented. Hypothetical thinking is plausible 'maybe thinking'. P. 85
Hypothetical thinking (confounding)	Hypothetical thinking involves identifying which confounders are plausible and estimating which are the biggest. P. 69
Index values exclusive	Index values can be exclusive (non-overlapping) or non-exclusive. Exclusive index values limit each subject to only one cell. Non-exclusive index values allow overlapping categories. P. 189
Index values exhaustive	Index values can be exhaustive or non-exhaustive. Exhaustive index values cover all relevant values of the index variable. Non-exhaustive index values omit some relevant values. P. 191
Incidence	Incidence **(**rate**):** a relative frequency: events per group size per unit time (2020 birth rate per 1,000 population). P. 199
Indexes	Indexes are the words in a table that indicate the content of the rows or columns.165
Inductive reasoning	Inductive reasoning with statements: reasoning where the conclusion is likely – but not certain – to be true even if all the premises are true. Often called informal reasoning. E.g., All swans I have seen are white. Therefore, all swans are white. P. 24
Inverse ratio	Inverse: a ratio in which the numerator and denominator are switched. In a part-whole ratio, the part and whole are switched. P. 275
Law of averages	The Law of Averages holds that as sample sizes increase, the sample averages will approach the population average. P. 338
Law of very large numbers	Law of Very Large Numbers. Qualitatively**:** The unlikely is almost certain given enough tries. Quantitatively: If an event has one chance in N and there are N tries, then one event is expected: the chance of at least one event is more likely than not. P. 92
Lift ratio	See the Over-involvement ratio.

Line: Dashed	Dashed lines indicate an association. P. 28
Line: Solid	Solid lines with arrows represent causation. P. 28
Longitudinal study	A longitudinal study is an observational study that involves repeated measures: measurement of the outcome at two or more different times on the same or similar subjects. P. 72
Margin cells	Margin cells are cells at the edge of a table (top, bottom, left or right) that include part or all the table. P. 170
Margin values	Margin values are the values at the edge of a table (indicated by "All" or "Total") that include all the subjects in a column or row. P. 170
Margin values	Margin values are either sums or averages. A sum is always bigger than the biggest value it includes, and an average is always smaller. P. 170
Margin-value rule	Margin Value Rule: If a margin value is a sum, then each component of the group is a separate part. If not a sum (if an average), then each component of the group is a separate whole. P. 184
Margin of error	Margin of Error: A range either side of a sample statistic that includes the population statistic 95% of the time. P. 97
Mean	The mean or average is the sum of the values divided by their count. P. 122
Measurement bias	Measurement (researcher) bias is systematic error arising from bad measurements, bad questions or bad judgements. P. 43
Mechanism	A mechanism is the means by which the predictor causes the result. Technically, a mechanism is a third factor that is caused by the predictor and causes the result. P. 55
Median	The median of a distribution is the middle value in a sorted list if odd (the middle of the two center values if even). P. 122
Medical test	Medical tests are tests that involve subjects with one of two conditions, that return one of two outcomes (positive or negative), and that are calibrated by how well the test confirms the known presence and absence of the condition. P. 277
Mix Matching	Mix-matching changes the mix in one group by applying a standard mix and recalculating the weighted average. P. 139
Mixed-fruit comparison	Mixed fruit comparison: a crude comparison of groups with different mixtures of a related variable. Also known as an 'apples and oranges' comparison. P. 136

Mode	The mode is the value or category with the highest frequency. 123 A unimodal distribution has just one peak. A bimodal distribution has two peaks. P. 123
Named Ratio	Named ratios are words that indicate the presence of a ratio: nouns such as rate, percentage or chance, or adverbs such as likely or prevalent in comparisons. P. 162
Necessary condition	A necessary condition is one that must be satisfied before a result can occur. (If the result occurs, the necessary condition must have been true.) P. 153
Non-response bias	Non-response bias: an outcome that is influenced by the mix of subjects that do not participate. P. 99
Nullify	Nullify is where an association vanishes after controlling for a confounder. P. 136
Observational study	An observational study is a study where the investigator observes the result of an exposure: the researcher has no control of anything. P. 71
Outcome	The outcome (result, response or dependent factor) in an association is the factor whose size or existence is being predicted, explained or influenced. P.16
Over-Involvement	Over-involvement ratio (RP) is a ratio of exposure percentages: the percentage of cases that were exposed divided by the percentage of non-cases that were exposed. If RP>1, then RR>1. P. 269
Part	Part designates the group (e.g., men) which if applied to the whole (e.g., soldiers) gives the part within that whole (e.g., male soldiers). P. 166
Part-whole percentage	A part-whole percentage gives the size of the part as a percentage of the whole. P. 166
Percent grammar	Percent grammar describes a part-whole ratio when there is no other named ratio keyword and the % symbol (or 'percent') is followed by 'of' or 'are'. P. 166
Percentage attributable to exposure	The percentage of cases in the exposure group that are attributable to the exposure is the excess between the exposure and control rates as a percentage of the exposure rate. P. 242
Percentage explained by confounder	The percentage difference explained by a confounder is the percent reduction in the original difference after taking into account the influence of the confounder. P. 150
Percentage grammar	Percentage grammar describes part-whole ratios using the keywords *percentage,* fraction *or proportion.* P. 176

Percentage points	Percentage points measure the difference between two percentages. P. 65
Percentile	A percentile is the percentage of subjects who have scores at or below a value. P. 119
Percentile pts	See Percentiles
Percentiles	Percentile points measure the difference between two percentiles. P. 120
Placebo effect	The Placebo Effect: subjects feel or do better after receiving a placebo: an inert (fake) medication or procedure. P. 98
Point	The point is a more disputable claim supported by evidence in an argument. P. 51
Political poll	Political polls forecast the vote in a political event: an election or in passing a proposition). P. 328
Population	A population is any group of interest. E.g., people, plants or stars. P. 94
Prediction vs explanation	Prediction (statistical): The item of interest is the part in a part-whole ratio. Explanation (statistical): The item of interest is a whole in a part-whole ratio. P. 283
Predictor	The predictor (explanatory or independent factor) in an association is the factor that predicts, explains or influences the existence or size of the outcome. P. 16
Prevalence	Prevalence (rate): the ratio of two counts: group count divided by population count at the same time (unemployment rate). P. 199
Prevarication	Prevarication is telling a half-truth by leaving out a most important part. P. 42
Probabilistic causation	Probabilistic causation is causation where ‘Something makes a difference sometimes’. P. 18
Prosecutor's fallacy	The prosecutor's fallacy (the base-rate fallacy) involves treating the confirmation accuracy of a test as though it measured the predictive accuracy thereby ignoring a base rate: the prevalence of the disease in the group. P. 278
Quasi-experiment	A quasi-experiment is an experiment that is not a scientific experiment or a clinical (randomized controlled) trial. Intervention may be done by humans or nature. P. 72
Question bias	Question bias: Wording questions in order to obtain a particular answer. P. 99
Quintiles	Quintiles are fifths of the group; quartiles are fourths and deciles are tenths. P. 121

Random assignment	Random assignment: randomly assigning subjects to the treatment and non-treatment (control) groups – or randomly assigning the treatment and placebo to each subject. P. 73.
Random samples	Random samples are samples in which some element of random selection or assignment was involved. P. 95.
Random sampling: benefit and cost	Random sampling is more likely to yield a representative sample on unknown factors than any other non-random process and it allows one to make statistical inferences, but it requires more time and money. P. 95
Randomized trial	Clinical trials are experiments involving random assignment. P. 73
Randomness	Randomness – pure chance – is the absence of any pattern that will help in predicting the next outcome. P. 37
Rank	Ranks (1st, 2nd, 3rd) measure the order or place of a value in a group of values with 1st being the best. P. 118
Rate	A rate is a ratio that uses *per* to introduce the unit of measure. 199
Rates: Four Kinds	Rates come in four kinds: frequency, prevalence, incidence and growth. P. 199
Rate grammar: clause-based	Clause rate grammar describes a rate using an entire clause: a verb separates the part and whole. P. 202
Rate grammar: phrase-based	Phrase rate grammar describes a rate using just phrases. P. 202
Rel. frequency distribution	A relative frequency distribution shows the percentage in each group by the column height or row length. P. 116
Relative pronouns	Relative pronouns (*who, that* and *which* as well as *what*, *where* or *when*) introduce relative clauses. P. 177
Relative risk	Relative risk (RR): the risk of an outcome in an exposure group (Re) divided by the same risk in the control group (Rc) written as "the relative risk of <outcome> for <exposed> is Re/Rc". P. 268
Replication	Repetition occurs when the experiment is repeated on the same subject in the same condition. P. 18
Representative samples	Representative samples are samples in which the sample matches the population on the relevant factors. P. 95
Representative sampling: benefit	Representative samples take less time and money than random samples, but the sample statistic is not necessarily the best predictor of the population statistic and the margin of error is unknown. P. 95

Researcher bias	Researcher bias is a change in outcome due to a researcher's knowledge of who is in which group. P. 99
Reversal	Reversal is where an association changes direction – a difference changes sign—after controlling for a confounder. P. 136
Safety effect	A Safety Effect is an increase in risky behavior because the subject knows they have safer equipment. P. 98
Sample	A sample is any part of the population. P. 94
Sampling bias	Selection (sampling) bias is systematic error [in the outcome] due to a non-representative selection from a population. P. 43
Sample error	Sample error is the actual difference between a sample statistic and the associated population statistic for a particular sample. P. 94
Sampling error	Sampling error is the *expected* difference between a *randomly-sampled* statistic and the associated population statistic. P. 95
Scientific experiment	A scientific experiment is an experiment that can be repeated. P. 18
Simple test	A simple medical test is one where the accuracy (the error) in confirming is the same for diseased and disease-free.. P. 279
Simpson's Paradox	Simpson's paradox is when an association has one direction at the group level and the opposite direction in each subgroup. P. 137
Single-blind study	A single-blind study uses a placebo to blind the subjects as to whether they are in the treatment or control group and eliminates the placebo effect. P. 100
Skewed distribution	A skewed distribution has one peak with one tail longer than the other. A right-skewed distribution is pulled to the right with a longer right tail. P. 117
Solid lines	Solid lines with arrows represent causation. P. 28
Specification	Specification applies an association in a group to a specific member of that group.52
Spurious association	A spurious association is one that vanishes (no difference, no correlation) after controlling for another factor. P. 135
Standard deviation	Standard deviation, the most common measure of spread, is related to the average variability of the data around the mean ignoring the sign. P. 156
Standardizing	Standardizing involves giving each group the same mixture of a confounder: either the mixture of the entire group or the mixture of one group chosen as the standard. P. 138

Statistical Literacy	Statistical literacy studies how statistics are constructed and manipulated. P. 15 Statistical literacy is the ability to read and interpret statistics in the everyday media. Statistical literacy is critical thinking about statistics as evidence in arguments. P. 2
Statistically significant	Statistically significant describes an outcome that is very unlikely if due just to chance. P. 96
Statistics	*Statistics are numbers in context*. P. 11
Stereotype	A stereotype is a judgment about all those in a group based on a characteristic of some members in that group. P. 52
Studies	Studies are either experiments or observational studies. Experiments involve doing (researchers treat subjects); observational studies involve only seeing (subjects are exposed). P. 71
Subgroup margin of error	The subgroup margin of error is the *maximum* error expected for a particular subgroup. P. 332
Subject bias	Subject (respondent) bias is when people lie or misremember in a systematic way. P. 43
Sufficient condition	A sufficient condition is one such that the result must occur if the sufficient condition is true. P. 153
Survey margin of error	The survey margin of error is the *maximum* error *expected* for statistics involving the entire survey. P. 332
Symmetric distribution	A symmetric distribution has a mirror reflection around its center. P. 117
Tables	Tables are organizations of data into cells that are arranged in rows and columns. P. 165
Take CARE	Take CARE is a good admonition in dealing with statistics. In this course, each letter stands for a kind of influence: C for Confounding, A for Assembly or Assumptions, R for Randomness and E for Error/Bias. P. 14
Take into account	See Control For.
Test and base	In a two-group comparison, the test (T) is the value being compared; the base (B) is the basis of the comparison. P. 62
Test part	A test-part is a test in a comparison and a part in a part-whole ratio. P. 246
Test whole	A test-whole is a test in a comparison and a whole in a part-whole ratio. P. 246

Three-centers rule	Three centers rule: The three measures of center have a natural order in most skewed distributions. It is alphabetic in English: mean, median and mode if skewed left; mode, median and mean if skewed right. P. 124
Times-based ratios	Times-based ratios are comparisons that involve division: times ratio, percent difference and times more. P. 66
Times-less comparison	The new times-less comparison, T is (B/T) times less than B, is unambiguous provided both Test and Base are always positive. P. 66
Treatment group	The treatment group is the group in an experiment that is treated. P. 70
Triangle diagrams	Triangle diagrams show the relationships between three related factors: a predictor, an outcome (result) and a related factor such as a confounder. P. 28
Uncontrolled study	An uncontrolled study involves just a single group. P. 70
Voter expectation	Who do you think will win the upcoming election? P. 329
Voter intention	If the election were tomorrow, who would you vote for? P. 329
Weasel words	Weasel words are words or phrases that suck the meaning out of a claim (just like weasels supposedly suck the yolk out of eggs). P. 42
Weighted average	Weighted average: weights the subgroup averages by their size. Arithmetically it first multiplies the average for each subgroup by the number in the subgroup, sums all these products and then divides the result by the total number in the group. P. 138

Index

100%
  column table 170
  row table 170
  total table 170
100% sum rule 172
A=Assembly 34
ABC association
  not causation 23
ABC words 20
  association 21
  between 21
  causation 20
accuracy
  confirmation 279
  prediction 279
alternate explanation 57
ambiguous words 20
anecdotal evidence 73
apples and oranges comparison .. 137
argument
  sound 106
  strong 106
  valid 106
  weak 106
arithmetic comparison
  template 63
arithmetic comparison, kind of
  1. simple difference 63
  2. times ratio 63
  3. percent difference 63
  4. times difference 63
assembly 34
  close reading 86
  comparisons 88
  definitions 87
  hypothetical thinking 87
  solutions 86
association 17
  crude 85, 106
  dashed lines 30
  definition 17
  flimsy 106
  graphical 19
  is not causation 23
  Simpson's Paradox 290
  solid 106
  spurious 292
  two-factor co-variation 17
  two-group comparison 17
  words 21
association of counts
  confounder decreases or increases 136
  confounder makes spurious 136
  confounder nullifies 136
  confounder reverses 135
attributable 242
  cases 244
attributed 242
availability bias 100
averages
  weighted 139
base 63
base indicators 63
base part 247
base rate fallacy 277, 279
base whole 247
Bayes comparison 270
Bayesian hypothesis testing
  50% Alternate 327
  Very Unlikely Alternate 328
between words 21
bias
  1-subject 45
  2-measurement 45
  3-selection 45
  availability 100
  comparison 100
  completion 100
  confirmation 100
  non-response 100
  researcher 100
  selection 100
big data 93
C=Confounding 29
CARE

  1Confounding 29
  2Assembly 34
  3Randomness 39
  4Error 44
cases
  percentage attributable 243
cases attributable 244
cases attributed
  #1 cases exposed 245
  #2 population exposed 271
  #3 population 271
causation 19, 26
  modals 243
  necessary 19
  probabilistic 19
  solid lines 30
  sufficient 19
  words 20
center and skewness 125
centers rule 125
chance 39
  evolution 338
chance grammar
  clause template 210
  phrase template 210
change, explain 70
clause description
  rates 207
clinical trial 73
close reading 86
coincidence 39
columns 166
common cause 30
common-part compare
  1 common part 247
  2 test whole 247
  3 base whole 247
common-part comparison 246
comparison
  arithmetic 63
  assembly 88
  base indicators 63
  benefit of 301
  complements 90
  crude 32
  mixed-fruit 137
  percentages 66
  times-less 67
comparison bias 100
comparison of ratios
  common-part 246
  distinct-part 246
  likely 258
comparisons, type of
  arithmetic 63
  ordinal 61
completion bias 100
components 172
confirmation accuracy 279
confirmation bias 100
confounded 29
confounder 30
confounder solutions 70
confounding
  averages 139
  counts/totals 135
  proportions 291
confusion of the inverse 276
context
  definition 7
control
  effect size 80
  study design 71
control for 83
control for, counts
  increase or decrease 136
  nullifies 136
  reverse 135
control of 83
controlled study 71
convenience sample 96
Covid death rates state compare . 274
Crimes
  gender 275
  race 275
cross-level inference 25
cross-sectional study 73
crude association 85
crude comparison 32
dashed lines 30

data distribution
  shape ... 117
definitions ... 87
denominator
  diabolical ... 189, 273
  missing ... 256, 264
Denominator
  missing ... 275
determiners ... 205
diabolical denominator ... 274
difference, explain ... 70
distinct part comparison ... 246
distinct-part compare ... 247
  1 test part ... 247
  2 base part ... 247
  3 common whole ... 247
distribution
  percentile ... 121
distribution grammar ... 183
double blind ... 101
drug effectiveness ... 269
E = Error/Bias ... 44
effect size ... 80
evaluation ... 26
Event ... 189
evidence
  anecdotal ... 73
  strength of ... 27
evolution
  chance ... 338
  inheritance ... 338
exclusive ... 189
experiment ... 72
  clinical trial ... 73
  doing ... 19
  double blind ... 101
  quasi ... 73
  randomized ... 73
  scientific ... 19, 73
  single blind ... 101
explanation
  alternate ... 57
explanatory power ... 301
exposure rate ... 243
factor ... 17
fallacy
  base rate ... 277, 279
  prosecutors ... 279
farther vs. further ... 65
frequency ... 200
frequency distribution ... 117
grammar
  rate ... 205
grammar rules
  likely comparison ... 257
  short form comparison ... 255
graphs, ratio comparison
  1 common parts ... 247
  2 distinct-parts ... 247
  3 common and distinct-parts .. 247
group ... 172
growth rate ... 200
half tables ... 184
  rates ... 203
  two-way ... 187, 252
halo effect ... 100
Hawthorne effect ... 99
hypothetical thinking ... 70, 87
incidence ... 200
inferences ... 54
inverse ... 276
journalistically significant ... 245
keywords
  ABC keywords ... 20
  association vs causation ... 20
law of averages ... 339
  compensation ... 339
  dilution ... 339
Law of Very Large Numbers ... 93
less vs. fewer ... 65
Lift ratio ... 270
likely ... 258
likely compare
  common part ... 260
  distinct part ... 260
likely comparison
  common- part ... 260
  distinct-part ... 261
lives saved attributable ... 270
longitudinal study ... 73, 78

margin ... 333
margin value rule ... 185
margin values ... 172
mean ... 123
measurement bias
  halo effect ... 100
  questions ... 100
  researcher ... 100
mechanism ... 57
median ... 123
missing margin rule ... 189
Missing-margin tables ... 189
mix matching
  arithmetic ... 140
mixed-fruit comparison ... 137
modals ... 243
mode ... 124
more important variable ... 301
movie study ... 73
multiple half tables ... 186
necessary condition ... 154
non-response bias ... 100
non-response rate
  phones ... 329
nullify ... 137
nullify or spurious association 136
observational study ... 72
ordered comparison of
  ratios ... 213
Over-involvement ratio ... 270
part-whole percentage ... 167
part-whole ratios, comparison
  1 common-part ... 247
  2 distinct-parts ... 247
percent difference ... 63
percentage
  part-whole ... 167
  syntax ... 177
percentage diff. explained by ... 151
percentages - compare
  two-way half tables ... 253
percentile ... 120
percentile score ... 120
phrase description
  rates ... 205
placebo effect ... 99, 101
political polls ... 329
polls, political ... 329
population ... 95
prediction accuracy ... 279
prediction versus confirmation ... 279
prediction vs. explanation ... 284
prevalence ... 200
prevalent ... 258
probabilistic causation ... 19
prosecutor's fallacy ... 279
quantitative data
  three key properties ... 117
quasi experiment ... 73
question bias ... 100
quintile ... 122
R = Randomness ... 39
random sample ... 96
random samples
  classification ... 332
  evaluation ... 332
random sampling
  classification ... 331
randomized experiment ... 73
randomness ... 39
rank ... 119
  calculate ... 119
rate ... 200
rate grammar
  clause ... 203
  clause template ... 207
  phrase ... 203
  phrase template ... 204
rates
  1frequency ... 200
  2prevalence ... 200
  3incidence ... 200
  4growth ... 200
  clause description ... 207
  common errors ... 206
  creating ... 202
  kinds ... 200
  not part-whole ... 206
  part-whole indicators ... 205
  phrase description ... 205

phrase-description 205
ratio
confusion of the inverse 276
inverse.................................... 276
ratio comparison graphic
1 common-parts...................... 247
2 distinct-parts 247
3 common and distinct-parts .. 247
ratio comparisons
other 66
ratio descriptions
ratio grammar 212
regression.................................. 315
relative pronouns 177
relative risk 269
repetition...................................... 20
representative sample 95
researcher bias 100
respondent bias
Hawthorne effect 99
placebo effect 99
safety effect 99
reversal 137
reversal condition 153
rows 166
safety effect bias........................... 99
sample.. 95
convenience 95
random 96
representative........................... 96
sampling bias vs. error 100
sampling error vs. bias 100
scientific experiment 73
Scientific experiment 19
selection bias 100
significant
journalistically 245
simple difference 63
simple random sample 331
Simpson's Paradox............. 138, 290
Simpson's paradox...................... 140
single blind 101
skewness
left vs. right............................ 119
rules for.................................. 125
snapshot study 73
solid lines 30
solutions
assembly 86
effect size 80
spurious association..... 136, 156, 292
standardizing
averages 139
statistical generalization 54
statistical literacy 9
statistical significance
applicable.............................. 337
assembly 324
Bayesian inverse 327
bias 325
confounding 322
flawed.................................... 320
influences 322
Unlikely Alternate 328
statistics
jewels 13
math definition 12
statistics contributions
random sampling 96
status 189
stereotype 54
study
anecdotal 73
controlled 71
cross-sectional 73
design 71
longitudinal 73, 78
movie 73
snapshot 73
study design 71
sufficient condition..................... 154
survey... 333
margin of error....................... 334
syntax
percentage 177
table
100% columns 170
100% rows 170
100% total 170
tables

columns 166
half.. 184
missing margin 189
multiple half 186
rows 166
two-way half.......................... 188
Take CARE
confounding............................. 29
mechanism............................... 57
taking into account.... See control for
template, ratio comparisons
likely common- part 260
likely distinct-part 261
percentage common-part........ 254
percentage compare, common-part 254
percentage compare, distinct-part 254
template, ratio descriptions
chance clause grammar 210
chance phrase grammar.......... 210
percent grammar.................... 168
percentage grammar 178
rate clause grammar 207
rate phrase grammar.............. 204
templates
arithmetic comparisons 63
ordered comparisons 62
templates, co-variation 69
test.. 63
test accuracy
confirmation 279
prediction 279
test part 247
test whole.................................... 247
three centers rule 125
times difference 63
times ratio 63
times-less
exception................................. 67
new.. 67
triangle diagram............................ 30
triangle diagrams 30
two-way half tables 188
uncontrolled study 71
variable 17
Very Large Numbers, Law of 93
weighted average........................ 139
weighted average line................. 143
whole
inappropriate.......................... 310
omitted 308

Review of Named Ratios

1. 'Among' and 'per' always introduce a whole.
2. Leading prepositions introduce, determine or delimit a common whole.
3. In statements, a single relative clause after "percentage" always contains a part.
4. Modifiers (leading adjectives or trailing phrases/clauses) typically take on the status of whatever they modify. (Note: one exception in the prior Percentage rule)

"Percent" Grammar (P. 166). Determiners and modifiers can be added.

"% of" present	"Among" absent	#% of <whole> are <part>.
"% of" absent	"Among" present	Among <whole>, #% are <part>.
"% of" present	"Among" present	Among <whole>, #% of <whole> are <part>

E.g., In the U.S., among women, 25% are smokers (or "25% smoke").
The main verb separates part and whole. The whole is on the same side of the main verb as the % symbol; the part is on the opposite side (or is the verb).

"Percentage" Grammar (P. 177). Other modifiers can be added.

"Percentage who*."	The percentage of__	who* are _____	is ___%
Among is absent	{whole}	{part}	. ##
"Percentage who*"	Among _____,	the percentage of ____	is ___ %
is absent	{whole}	{part} .	##
"Percentage who*",	Among ______, the	percentage of __ who* are __	is ___ %
of and among	{whole}	{whole} {part .	##

* Other relative pronouns include *that, which, what, when* and *where*.

Rules for Decoding Tables of Ratios (Percentages or Rates)

Margin values are either sums or averages. A sum is always bigger than the biggest value it includes, and an average is always smaller. P. 170

100% Sum Rule: If a margin value is a 100% sum, then the group is whole and each component is a part. P 171

Margin Value Rule: If a margin value is a sum, then each component of the group is a separate part. If not a sum (if an average), then each component of the group is a separate whole. P. 184

Missing Margin Rule: If margins are missing and the index values are exclusive, they are wholes (unless they add to 100%). P. 188

Questions: Percent versus Percentage Grammar (P. 204).

What percentage of <whole> are <part>?
What is the percentage of <whole> who are <part>?

Comparing two numbers. One is test (T), other is base (B).

1) Difference: # = (T - B): _______ is ______ more/less+ than ______.
{test} # or # percentage pts {base}
2) Times ratio: # = (T/B): ______ is ___ times as much/many as _____
{test} # {base}
3) Percent difference: # = 100(T-B)/B ___ is __ % more / less+ than ___
4) Times difference: # = (T-B)/B ___ is __ times more/less+ than ___
{test} # {base}
+: Difference comparisons allow "er" endings: greater, smaller, etc.
'Often' and 'frequently' can be used with the three ratio comparisons

COMPARING RATIOS: Common Part p. 245. [Distinct part: p. 245]

To delimit a common whole, leading phrases can be added before these templates.
These templates show ratio and percent difference. Use templates above for others

"Percentage" Grammar, Long-Form Compare (P. 253)

The percentage of _________	that are _______	is ___	times as much as	the percentage of _______	that are _______
test whole	common part	##	compare	base whole	common part

Percentage of gals who run is _ times as much as the percentage of guys who run

The percentage of ________	among _______	is ___	% or times more than	the percentage of _______	among _______
common part	test whole	##	compare	common part	base whole

Percentage of runners among gals is ____ % more than that percentage among guys

"Likely" Grammar Rules: Common part p. 257; [Distinct Part p. 260]

1. "among" always indicates a whole
2. "to" indicates a part. (Also, to be, to do, to have, etc.)
3. A part-whole compare must have at least 3 part-whole terms with at least one part and one whole.
4. "as X is" or "than X is" means X is *linked* to the subject. Two linked terms have the same part-whole status.
5. "is likely to" without an object (e.g., *is likely to occur* or *is likely to happen*) indicates the subject is the part.

Common Part Compare: "Likely Among". Part as subject (P. 259)

______ is/are	___	times as likely	among/in _____	as	among/in _____
common part	##	compare	test whole	Indicate	base whole

In 2019, U.S. 12^{th} graders were twice [two times] as likely to smoke as 8^{th} graders.

Common Part Compare: "Likely To". Whole as subject (P. 259)

______ is/are	___	% more/less likely	to _________	than	is/are _______
test whole	##	compare	common part	indicate	base whole

E.g., In 2000, women were 25% more likely to smoke than [were] men