STATISTICAL LITERACY: CRITICAL THINKING ABOUT STATISTICS

Milo Schield
University of New Mexico
SchieldMilo@UNM.edu

–Many – if not most – arguments today involve statistics: social statistics. These statistics are quite different from mathematical numbers. Statistics are socially constructed by people with goals and motives. And since they are based on reality they can be influenced. This presentation introduces statistical literacy as a new discipline and as a course (Math 1300) taught at the University of New Mexico. The focus is on introducing those ideas that students need to decode and evaluate the statistics in the everyday media, in tables and in graphs. Statistical literacy, quantitative rhetoric, is argued to be a necessary skill in order to deal with data based arguments in modern society.

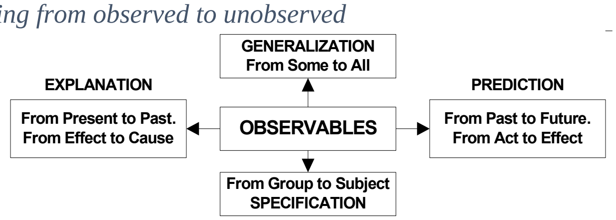
INTRODUCTION:
Critical thinking about statistics? I'm sure you are wondering. Critical thinking in mathematics? Isn't that the last place we'd go to look for critical thinking? Mathematics is mainly deductive. Deduction is barely critical thinking. But stay with me. Let's see how it turns out.

CRITICAL THINKING:
Dr. Elder gave a talk last fall pointing out that critical thinking faced some headwinds in being adopted in the academy. One of those headwinds was philosophy. I agree, but I want to go on and say that the social sciences, in their own way, shun critical thinking – as does traditional Statistics.

Why does Philosophy have this problem with critical thinking? Because Philosophy was unable to solve or resolve the problem of induction. Induction here is not about motors or inducing labor at birth. In philosophy, induction is reasoning from evidence.

In Figure 1, we have what we can see or observe in the center. At the edges are things that we don't generally see. The outward arrows show the mental movement from observed to unobserved.

Figure 1. Induction: Reasoning from observed to unobserved



- On the right side is prediction: the movement from the past to the future: from act to effect.
- On top we have generalization: the movement from some to all: from sample to population.
- On the bottom we have specification: the movement from the properties or behavior of a group to the properties or behavior of a member of that group. Suppose that cars that make a certain sound have a certain mechanical problem. Suppose that your car makes that same sound. We can infer that your car has the same mechanical problem.
- On the left we have explanation: the movement from present to past (from effect to cause). For example, why did something happen? How did something happen?

These are all inductions: reasoning from observed to what is unobserved. These inductions are what are responsible for human progress. Consider this example:

When a child is about six months they often enjoy playing the 'peek-a-boo" game. The adult covers their face with a pillow, pauses and then suddenly shows their face to the child. Often the child responds with amazement and laughter. After a while, they lose interest in the game.

What is happening? As adults we can't be sure. But it seems that the child realizes that they can see part of you: your body. But they can't see all of you: your face. And they conclude that your face still exists even though they can't see it. This is a tremendous accomplishment: an incredible mental leap. For me, this is the first sign of human intelligence in the young child. This is induction: to reason from what is observed to what is unobserved.

There was extensive human progress during the scientific revolution with Copernicus, Galileo and Kepler. But in 1748, a philosopher, David Hume, argued that we cannot generalize with certainty. This is the problem of induction. A later philosopher, Emanuel Kant, stated that Hume's claim posed "a most fundamental challenge to all human knowledge claims." Another philosopher, C. F. Broad, stated that "induction is the glory of science and the scandal of philosophy.'

To put it most bluntly, one might hear Hume saying, "If it is not empirical (if can't count it, touch it or see it), it's 'crap'." If true, this claim eliminates most of the things we find interesting. What is true? What is good or right? What is beautiful? Philosophy can't really argue convincingly on such matters. This is why Philosophy can't really engage in critical thinking.

This failure of the humanities to deal with the problem of induction led critical thinking to be (Bardi[1])

- Reductionist: 'explaining' complex phenomena in terms of more elemental components.
- Positivist: limiting the 'real' to what is physically observable or which can be proved.
- Quantitative: understanding qualities in terms of measurable qualities.

This failure of the humanities to deal with the problem of induction led to the rise of psychology and the social sciences. These 'soft' sciences noted that the hard sciences avoided induction (generalizing) by performing experiments: repeatable experiments. And so, to be like the hard sciences, psychology and the social sciences focused on experiments or studies.

But with human subjects the social sciences could not conduct scientific experiments. Yes, they might repeat the experiment with the same subjects. But these subjects would not be in the same condition.

So psychology and the social sciences turned to statistics where one-off (one-try) experiments and studies were the norm. Unfortunately, statistics has tended to shun critical thinking. It did so by staying close to mathematics. And this mathematical approach to statistics has permeated psychology and the social sciences.

Statistics has shunned critical thinking in two ways.

- by *using a fixed level for statistical significance* – disregarding the plausibility of the conjecture. "Statistical significance" is a big concept in statistics. We are going to disregard it here except to say that statisticians used a fixed level regardless of the plausibility of the conjecture. The conjecture might be very plausible: putting a fertilizer on a field to improve the yield of a crop. The conjecture might be very implausible: testing humans for ESP – extra sensory perception. These are vastly different conjectures. But statisticians don't include that fact in their analysis of statistical significance.
- by *ignoring the influence of confounding in observational studies*. Almost all the studies in Sociology are observational studies: the researchers are passive; they just observe. The resulting associations are true, but they prevaricate. These crude association don't tell the whole story; they tell a half-truth. I will give an example of this later. Taking into account the influence of a confounder can change the size and the direction of an association. Students in Sociology seldom – if ever – learn this in their research methods course. Finally, those teaching introductory statistics ignore how controlling for a measured confounder can change statistical significance into insignificance (and vice versa).

CONFOUNDING:

Confounding is something that is 'found with' something. Confounding is something that confuses.

Confounding is the 'elephant' in observational studies. Statistical educators know this. But they don't teach it in the introductory course. They should teach it. But, there isn't time in teaching a traditional introductory statistics course – given all the topics they must teach to satisfy the statistical needs of those students that are required to take a statistics course.

Confounding is where the critical thinking comes in.

[1] Bardi, J. Thinking Critically about Critical Thinking www.personal.psu.edu/jfb9/essay2ThinkingCritically.html

SOCIAL STATISTICS
Today's students are interested in arguments. Most stories in the everyday media involve controversial claims. These asserting these claims give reasons – they provide arguments.

Many of these arguments involve statistics. Most of these statistics are social statistics: statistics involving people.

Many of these claims involve a certain kind of statistic: disparities in outcomes. Disparities in education, suspension and graduation. Disparities in policing, crime, arrests, sentencing and prison. Disparities in money: in income, wages, assets, loans and wealth. Disparities in health: in healthcare, homicides and death.

These disparities in outcomes typically involve groups: groups involving gender, race, ethnicity, socio-economic class, religion, politics, age, etc.

All of these outcome disparities by group are associations. These associations rely on statistics: social statistics. Almost all of these statistics are based on observational studies. They are crude associations: they are true, but they tell just a part of the story. In many – if not most – cases, there is a story behind each of these statistics.

STATISTICAL LITERACY AT THE UNIVERSITY OF NEW MEXICO
In fall 2021, the University of New Mexico offered students in non-quantitative majors a new catalog course: Math 1300 Statistical Literacy. In 2021-22, seven sections were offered with 15 to 30 students per section.

This catalog course (Math 1300) satisfies a mathematics requirement in the New Mexico general education curriculum. Here is the catalog description for Statistical Literacy:

> *Participants will study the social statistics encountered by consumers. Investigate the story behind the statistics. Study the influences on social statistics. Study the techniques used to control these influences. Strong influence on confounding.*

Consider these phrases:

- *Investigate the story behind the statistics.* In their 12 years of mathematics, students have never seen or heard of a story behind the numbers. They have investigated the story behind the story in their majors: history, journalism, political science, English, etc. They have gone beneath the surface and looked for other connections: other explanations. But in twelve years of mathematics, they have never done anything like that in dealing with a number.
- *Study the influences on social statistics.* In their 12 years of school mathematics, they have never investigated the influence on a number.
- *Study the techniques used to control these influences.* You don't do that in mathematics – at least not the mathematics they took in high school.
- *Strong influence on confounding.* Most introductory courses never mention the word 'confounding'. Most introductory course don't have 'confounding' or 'lurking variable' in their table of contents. Even if they do, it typically involves a way in which association is not causation. It then disappears from the course entirely.

Math 1300 is a very different course from the traditional introductory statistics course. This statistical literacy course has less than a 30% overlap with a traditional statistics course. Math 1300 is a combination of three topics: statistics, critical thinking and literacy. This course isn't a regular math course that uses symbols for variables. This course uses ordinary English to describe counts, amounts and ratios: rates and percentages. These words are often being used in very technical ways. Ways that students are not familiar with. But the words are just common ordinary English.

As a discipline, statistics studies variation. There are two kinds of variation: random and systematic.

- Random variation is studied in the typical introductory statistics course. Random variation is the variation in sample statistics in small samples that are randomly selected or randomly assigned. This course deals with statistical inference: topics such as margin of error, confidence intervals, a test of hypothesis, and statistical significance.
- Systematic variation is studied in Statistical Literacy. Systematic variation can be found in small samples and in big data. Topics include the various influences on a statistic such as confounding (study design),

assembly (how the statistics are defined, counted, measured, presented, etc.) and Error or bias. This material may be taught in the second statistics course: the regression or modeling course.

These two courses have two different audiences.

- Traditional statistics is designed for the *producers* of statistics: students in STEM majors (Science, Technology, Engineering and Mathematics).
- Statistical Literacy is designed for the *consumers* of statistics: students in the non-quantitative majors like journalism, political science, and history. Majors like archaeology, film, art and music. Majors like geology, epidemiology. These are majors that don't require a particular mathematics course.

So who determines the content of this new statistical literacy course? Is this just another algebra course in disguise? No. The teachers don't pick the content. The content is determined by the statistics in the everyday media. This is a different course – a very different course.

Dr. Joel Best cut to the heart of the matter when he said, "Statistics are socially constructed." Best is the author of a great book: "*Lies, Damned lies and Statistics*." He identified the social construction of statistics as a most important – the most fundamental – aspect of all reality-based statistics.

Statistics are more like words than numbers. Statistics are created by people: people with motives, values and goals. It is so easy to think of statistics as just being numbers. Yes, statistics are numbers. But they are numbers in context – numbers in reality. Statistics can be influenced by the context – by the reality.

In arithmetic, one plus one is always two. But in 'bunny math', one bunny plus one bunny can give more than two bunnies – under certain conditions. In 'ice-cube math', one ice-cube plus one ice-cube can give less than two ice cubes – under certain conditions. The reality matters. Statistics can be influenced by the context.

Take CARE:
What is the best advice to anyone who is dealing with statistics? *Take care*! Statistics can be influenced.

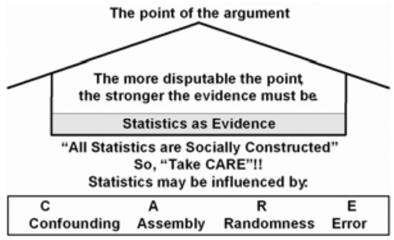
In Math 1300, all the influences on a statistic are grouped into four categories. The first letter in the name of each category matches the four letters in CARE.

- *C: Confounding*. Statistics are confounded (confused) by related factors. Critical thinking is required to think of what related factors are most likely to influence a given statistic.
- *A: Assembly or assumptions*: statistics are influenced by how they are defined, counted, measured, summarized, compared and presented. Consumers must use critical thinking to think of how a statistic could have been defined, counted, measured, etc.
- *R: Randomness*: Not just the randomness in sample statistics in small samples, but the randomness in extreme outcomes, or patterns.
- *E: Error or bias*. Error includes the confusion of the inverse: confusing the "percentage of women who smoke" with the "percentage of women among smokers". Bias includes subject bias, measurement bias or selection bias. Here are some examples: Subject bias: people understate their age or weight; they overstate their income. Measurement bias: Asking loaded questions. Selection bias: Popes and political leaders live longer than the average person. True, but people do not become pope or a political leader until they are already older. On average, older people live to an older age than do infants.

The last two topics (Randomness and Error/bias) are the basis for the traditional introductory statistics course. The first two (Confounding and Assembly) are the primary focus of a statistical literacy course. The difference is the weight. In the traditional course, randomness and error make up at least 70% of the course. In this confounder-based statistical literacy course, randomness and error/bias are less than 30%.

So, "take CARE" when dealing with social statistics. Critical thinking is required to analyze and evaluate these influences. A statistic that is very susceptible to being influenced provides weak support in an argument.
An argument is something mental: an arrangement of claims that support a more disputable claim. In order to understand arguments, here is a physical analog: a house.

Figure 2. An Argument is Like a House



In a house, the basement supports the walls; the walls support the roof and the point of the roof. In an argument, the reasons are like the basement and the walls: they support the point of the argument. In an argument, the point is a claim that is less observable and thus more disputable than the claims used as evidence.

The statistics are typically in the basement. As noted before, statistics are socially constructed so they can be influenced and manipulated. The best advice when dealing with statistics is "Take CARE". The influences are grouped into four categories where the first letters match the four letters in CARE as shown in Figure 2.

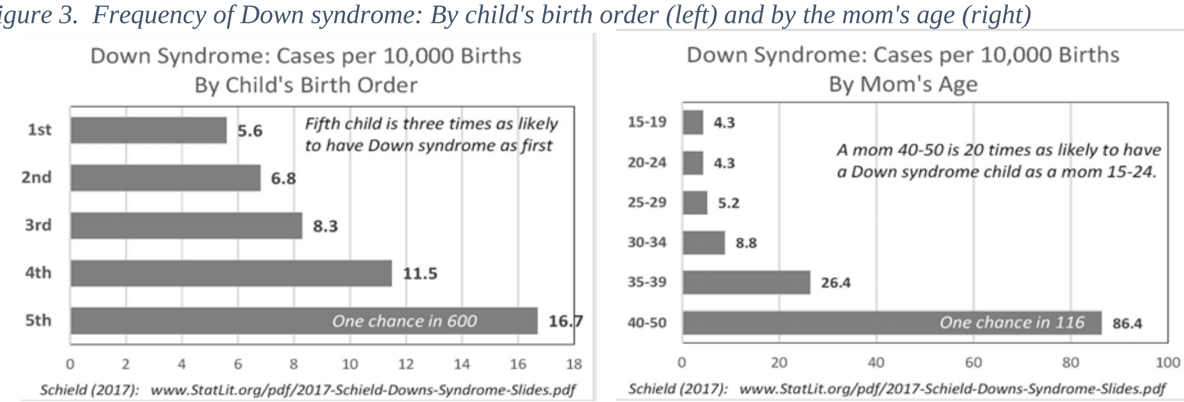
Students like "CARE". It gives them a structure. When given a list of 25 ideas or topics and asked to rank which idea they considered the most valuable, most students chose "Take CARE". Here are some more examples for each of the four letters of CARE:

C: Confounding

#1: To better understand confounding, consider statistics on a birth defect: Down syndrome. As shown on the left side of Figure 1, the later the child in the birth order, the more likely they are to have a Down syndrome. The fifth child is three times as likely to have Down syndrome as the first child.

This association is true – given this data. But it may be a crude association. If there are other differences between the first child and the fifth child, this comparison may be a mixed fruit comparison: an apples and oranges comparison.

Figure 3. Frequency of Down syndrome: By child's birth order (left) and by the mom's age (right)



What else could influence having a Down syndrome child? The mom's age. The right side of Figure 1 shows the same data organized by the mom's age.

As the mom's age increases, the chance increases that her child will have a Down syndrome. Moms age 40 to 50 are about 20 times as likely to have a Down syndrome child as are moms 15 to 24.

These two explanations are connected: confounded. The later the child in the birth order, the older the mom. Perhaps, the association we saw between Down syndrome and birth order is really spurious. Perhaps the only influence is the mom's age.

People who read home and fashion magazines are more likely to get pregnant than people who read car and sport magazines. We know that pregnancy isn't caused by magazines. We know that only women can get pregnant. We quickly recognize that women are more likely to read home and fashion magazines than men. QED. This association may be confounded by gender.

A: Assembly
Consider 'bullying' in middle school. Suppose bullying is defined as "using physical force to get what you want from others without their consent."

- How could we change the definition to *decrease* the quantity of bullying – without changing anyone's behavior? Simple! Change the definition. Suppose we require a wound, a visit to the Nurse's office, a visit to the hospital, or a death certificate. Each of these restricts the concept and thereby decreases the amount of bullying.
- How could we change the definition to *increase* the quantity of bullying? Simple! Change the definition. Re-define bullying to include gossip, slander and spreading false rumors. The more 'doorways' we allow for bullying (A or B or C), the greater the number of events involving bullying.

R: Randomness
Some statisticians use less than one chance in 20 as the condition for an association between two statistics to be statistically significant. Using a fixed rule (one chance in 20) violates the Carl Sagan rule: "Extra-ordinary claims require extra-ordinary evidence". This fixed-level rule has no way to take into account the context of the situation.

E: Error/Bias
Subject bias: People sometimes misremember or lie. Some sick people feel better after getting a placebo: a fake treatment. Measurement bias: Leading questions "Do you think our hard-working President is doing a good job in these extremely challenging times?" Selection bias: Unhappy customers are more likely to participate in surveys than happy customers.

AUXILIARY TOPICS:
Statistical Literacy (Math 1300) has a strong emphasis on hypothetical thinking, using ordinary English and writing.

Hypothetical Thinking
Confounding and assembly both require hypothetical thinking. Hypothetical thinking.

- Would having the first child at age 40 be more risky than having 5 kids before turning 30?
- How could a statistic have been defined when the definition isn't given?

Ordinary English

- Association is not … causation. *People who shave their face are taller than those who shave their legs.* We know that shaving does not influence height. What else does influence height? Gender. Men are more likely to shave their face; women are more likely to shave their legs.
- Disparity is not … discrimination. *Tall people are 10 times as likely to be in prison as short people.* Again, we are pretty certain that height is not connected to being in prison. What else influences being in prison? Gender! Men are much more likely to be in prison than women are. Men are generally taller than women. This disparity, a crude association, is confounded by gender.
- "The percentage of women who are runners" is different from "the percentage of women among runners."
- "Suicide is more likely among widows than [among] widowers is different from "Widows are more likely to commit suicide than widowers [are]."

Writing/Arguing
Writing is typically 20% of the grade in Math 1300. Students analyze and evaluate the statistics in news stories in the everyday media.

Statistical Literacy (Math 1300) is a new course. It has a new textbook. It has been field tested by over a thousand students and taught by more than a dozen teachers. Students at the University of New Mexico found Statistical Literacy (Math1300) valuable. Here is data from 76 students in an anonymous survey from the fall 2021 class.

Q1. *How valuable is this course in helping you read and interpret everyday statistics?*
Negative (1%), Neutral (4%), Some value (12%), Fair value (38%), Highly valuable (45%).

Q2. How helpful was this course in developing your critical thinking skills?
Not helpful (0), Neutral (5%), Somewhat (21%), Very (42%), Extremely (32%)

Most students (83%) found this Statistical Literacy course very valuable (fair value or highly valuable) in helping them "read and interpret everyday statistics". Most students (74%) found this Statistical Literacy course very valuable (very or extremely valuable) in "developing their critical thinking skills". (Schield, 2022a).

I challenge anyone to get results like these from any other course that satisfies a mathematics or statistics requirement in a college general education curriculum.

CONCLUSION

Students need to read and interpret social statistics in order to evaluate today's arguments. Faculty in the humanities need to persuade their mathematics' colleagues to offer statistical literacy for students in non-quantitative majors.

ACKNOWLEDGMENTS

Thanks to the Critical Thinking Foundation for accepting my proposed guest lecture for their 2022 Conference on Critical Thinking. Two lectures were prepared and recorded. This one involves critical thinking about statistics. It covers all the influences on a statistic and is more general. The other involves critical thinking about confounding. It has more mathematical content. It shows what it means to 'take something into account.'

BIBLIOGRAPHY:

Bardi, John (unknown). Copy at www.personal.psu.edu/jfb9/essay2ThinkingCritically.html

Schield, M and Raymond, R (2009). Distinguishing Association from Causation in Media Headlines. *ASA Proceedings of the Section on Statistical Education*. P 4371-4385. Copy at www.statlit.org/pdf/2009SchieldRaymondASA.pdf

Schield, M (2018). Confounding and Cornfield: Back to the Future. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward*. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July 2018), Kyoto, Japan. Copy at http://www.statlit.org/pdf/2018-Schield-ICOTS.pdf

Schield, M. (2021). The Diabolical Denominator. Mathfest. www.StatLit.org/pdf/2021-Schield-MathFest.pdf

Schield, M. (2022a). Statistical Literacy UNM Math 1300: First Year Results at Augsburg College. JSM 2022 *Proceedings of the Section on Statistical and Data Science Education*. P 1235-1265. Copy at www.statlit.org/pdf/2022-Schield-ASA.pdf

Schield, M. (2022b). Statistical Literacy: Critical Thinking about Statistics.
Slides at www.StatLit.org/pdf/2022-Schield-CCT2-Slides.pdf
Paper at www.StatLit.org/pdf/2022-Schield-CCT2.pdf
Video at https://drive.google.com/file/d/14QFCPm_WijQUZiU7IN9iCuYptVAwOHs3/view

Schield, M. (2022c). Statistical Literacy: Critical Thinking about Confounding.
Slides at www.StatLit.org/pdf/2022-Schield-CCT-Slides.pdf
Paper at www.StatLit.org/pdf/2022-Schield-CCT.pdf
Video at https://drive.google.com/file/d/1RCu7ZC81IWdQGR4RpU0LIhWbuFN0ggx4/view?usp=sharing