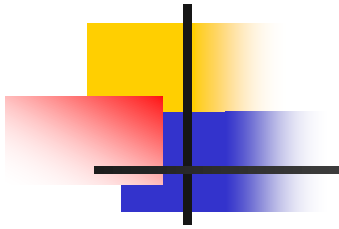
Data Science Ethics: A Checklist for Statistics Educators

Jessica Utts
Professor Emerita of Statistics
University of California, Irvine

A Caveat

- This talk is an abbreviated version of a previous talk and paper, but with some (controversial) new content as well.
- Previous talk (slides and video) at my home page under "Representative presentations"
  - https://www.ics.uci.edu/~jutts/
- Utts, Jessica (2021) "Enhancing Data Science Ethics Through Statistical Education and Practice," *International Statistical Review*, 89.1: 1-17.
- In this talk I will sometimes refer you to those sources for details.** Think of this talk as a sampler!

Outline



- Part 1: Setting the stage with some history
- Part 2: The frightening present
- Part 3: How statistics educators can help
- Part 4: Encouraging statistical literacy

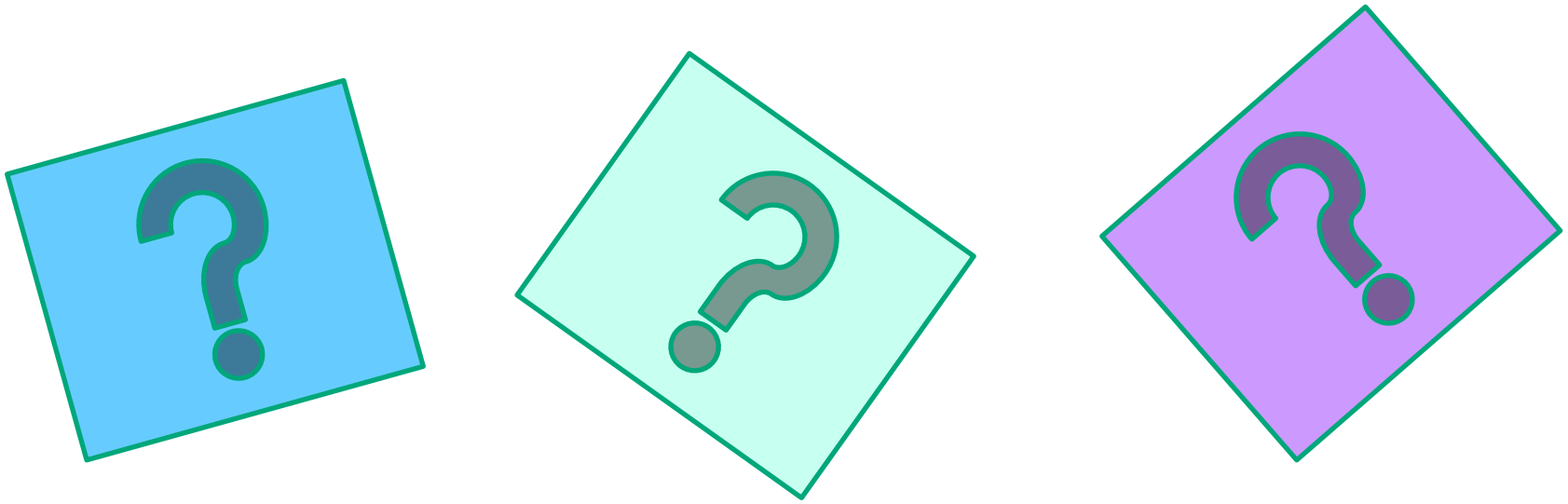
Part 1:
Setting the stage with some history

A story from my early teaching years

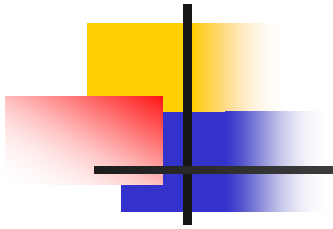
- Assignment in engineering class: design a pipeline to send blood from a poor developing nation to a rich developed one.
- The students got to work, discussing the optimal diameter for the pipe, how to go under a body of water, methods for keeping the blood fresh, etc.
- After watching awhile the professor told them they all failed, because not one of them had questioned the ethics of the assigned task.
- “This is a class in engineering not ethics,” the students protested

Moral of the Story

- Train students to ask *WHY* before asking *how*.
- Is the task ethical? Are there pros and cons?
- Who might benefit? Who might suffer?



Example: GPS Map program



Is it ethical…

- To clog roads by sending everyone on the same route when leaving a large event?
- To send cars through high-crime areas?
- To even identify high-crime areas?
- To send pedestrians through high-crime areas?
- To increase traffic in residential areas?
- What about school zones?

Part 2:
The frightening present

Why this topic? Why now?

Why this topic? Why now?

Why AI Must Be Ethical — And How We Make It So

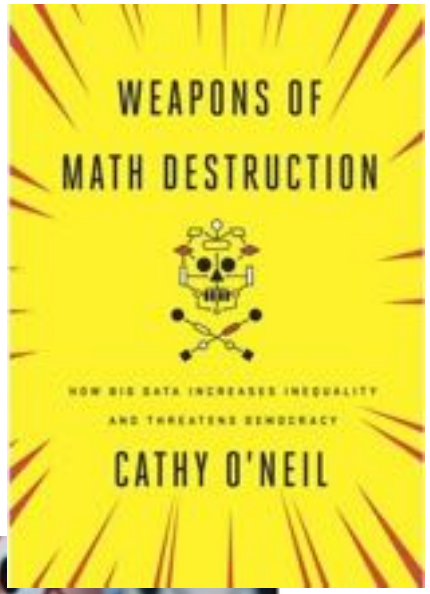
When artificial intelligence decides who's worthy, who's right, and who's criminal, you can only hope it makes the right call.

Jacob Bergdahl Follow

Jul 14 · 7 min read ★

It's 2021. The rain is pleasantly pouring outside. You're having an afternoon tea, while working on your next project on your laptop. Your workflow is interrupted by a phone call. It's Danielle from Subtling. You were there last Friday on a job interview that you're

Why this topic? Why now?



Why AI Must Be Ethical — And How We Make It So

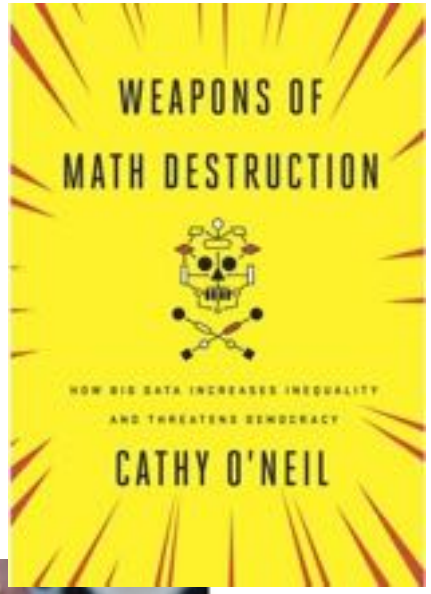
When artificial intelligence decides who's worthy, who's right, and who's criminal, you can only hope it makes the right call.

Jacob Bergdahl Follow

Jul 14 · 7 min read ★

It's 2021. The rain is pleasantly pouring outside. You're having an afternoon tea, while working on your next project on your laptop. Your workflow is interrupted by a phone call. It's Danielle from Subtling. You were there last Friday on a job interview that you're

Why this topic? Why now?



Why AI Must Be Ethical — And How We Make It So

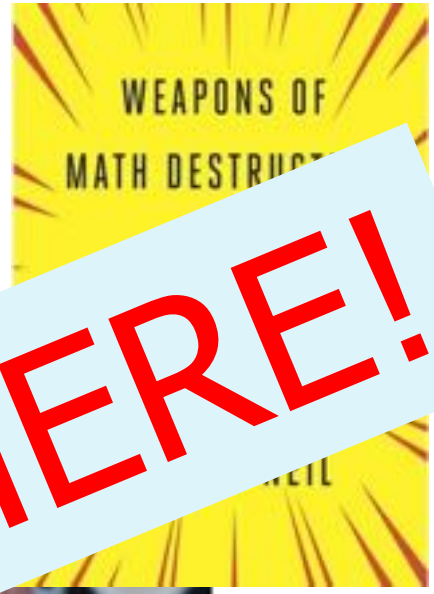
When artificial intelligence decides who's worthy, who's right, and who's criminal, you can only hope it makes the right call.

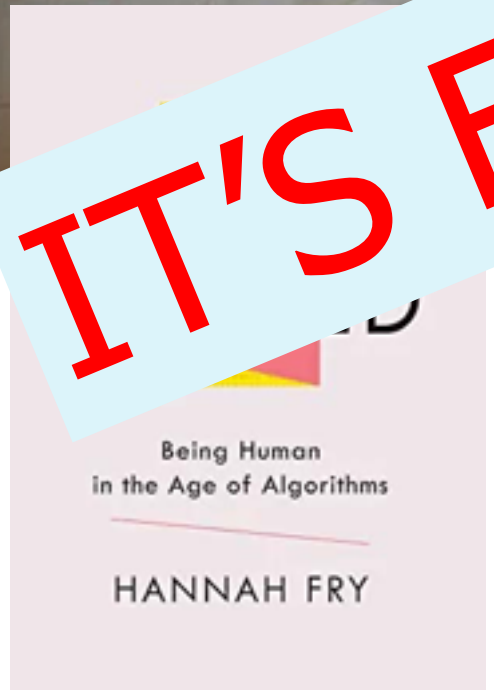
Jacob Bergdahl Follow

Jul 14 · 7 min read ★

It's 2021. The rain is pleasantly pouring outside. You're having an afternoon tea, while working on your next project on your laptop. Your workflow is interrupted by a phone call. It's Danielle from Subtling. You were there last Friday on a job interview that you're

Why this topic? Why now?





Why AI Must Be Ethical — And How We Make It So

When artificial intelligence decides who's worthy, who's right, and who's criminal, you can only hope it makes the right call.

Jacob Bergdahl Follow

Jul 14 · 7 min read ★

It's 2021. The rain is pleasantly pouring outside. You're having an afternoon tea, while working on your next project on your laptop. Your workflow is interrupted by a phone call. It's Danielle from Subtling. You were there last Friday on a job interview that you're

https://towardsdatascience.com/why-ai-must-be-ethical-and-how-we-make-it-so-952d0b1d1f

1/9

IT'S EVERYWHERE!

Why this topic? Why now? Continued…

- Lines are blurring for data science:

  Statistics/machine learning/artificial intelligence
- Our students' jobs reflect this cross-over
- Traditional ethical issues for statisticians
  - See for instance "ASA Ethical Guidelines" (April 14, 2018)
- Not enough. Complexity => new ethical issues
- Educate students on ethics of decisions/interpretations
  - As data scientists
  - As members of multidisciplinary teams
  - As consumers

Example: ACLU Congress Face Recognition Study**

- Facial recognition system Amazon offers to public (Rekognition), used default settings.
- Using database of over 25,000 arrest records, looked for matches with public photos of all Congress members
- Found 28 (out of 535 members of Congress) supposed "matches" to criminals; disproportionately members of color, including many from the Black Caucus.

Other data science ethics examples

- Bias in hiring algorithms from bias in training data.
- Algorithms used by judges to decide who is likely to commit (another) crime. Extended example in *ISI* paper; program called COMPAS.**
- Medical diagnostic algorithms trained on data excluding certain sub-populations.
- Using genealogy DNA databases to solve cold cases.**

A few examples from the media

The New York Times

She Was Arrested at 14. Then Her Photo Went to a Facial Recognition Database.

With little oversight, the N.Y.P.D. has been using powerful surveillance technology on photos of children and teenagers.

8/2/2019 DNA Website Had Unwitting Role in Golden State Manhunt - Bloomberg

Prognosis

DNA Website Had Unwitting Role in Golden State Manhunt

Kristen V. Brown
May 29, 2018, 11:34 AM PDT
Updated on May 29, 2018, 2:00 PM PDT

- GEDmatch, founded by a Florida grandfather, pools genetic data
- Privacy debate swirls after investigators sifted site's users

Curtis Rogers is used to helping people track down their relatives. The 79-year-old Florida grandfather of six founded a genealogy website that helps hobbyists like himself trace the branches on their family tree.

BUSINESS NEWS OCTOBER 9, 2018 / 8:32 PM / 10 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin 6 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

amazon

Very Recent Example: Alzheimer's Drug

FDA Approved Biogen Alzheimer's Drug Despite Some Staff Concerns

The agency has faced criticism for approving the drug, Aduhelm, without clear proof it works

By Joseph Walker and Thomas M. Burton, Updated June 22 [2021], 6:21 pm, EDT

"The U.S. Food and Drug Administration approved the first new Alzheimer's drug in decades over the objection of agency statisticians who said there was insufficient evidence to support approval, according to newly released internal memos."

What's the Controversy?


THE WALL STREET JOURNAL.

Jessica Utts

Id U.S. Politics Economy **Business** Tech Markets Opinion Life & Arts Real Estate WSJ. Magazine Sports

Latest Experimental Alzheimer's Drug Fails Testing

Drugmakers Biogen and Eisai ended studies of treatment, deeming it unlikely to benefit patients in latest research setback

March 21, 2019: "The search for new Alzheimer's disease treatments hit another big setback on Thursday when drug makers Biogen Inc. and Eisai Co. said they would terminate two late-stage studies of an experimental drug after determining it would likely fail to help patients."

Long story short…
There were two Phase 3 clinical trials

- Primary endpoint: Cognitive Dementia Rating Scale Sum of Boxes (CDR)
  - Normally, two studies with consistent results are required for drug approval, with at least one statistically significant.
  - Only one study (Study 302) showed statistical significance, and originally was terminated early for "futility".
  - The other study (301) showed results in the opposite direction for high dose groups; FDA asked the panel to ignore that one and use a preliminary study instead as the 2nd study.

Story, continued…

- *After* Study 302 was terminated (early, for futility) and was unblinded, Biogen continued to look at data
  - Evidence of reduction in beta amyloid for drug vs placebo
  - FDA approved the drug for 9 years based on that surrogate
  - But there is no convincing evidence that reduction in amyloid improves cognitive functioning
  - Evidence of better cognitive functioning for high dose group in Study 302, but Study 301 was in opposite direction

So What?

- Drug costs $56,000 a year (Medicare? Under review)
- Serious side effects (brain swelling)
- Only tested in patients in earliest stages of disease
- Because it was not effective in reducing *clinical* symptoms, how will doctors know if it's working, and when the patient should stop taking it? (Brain scan required to check for amyloid reduction.)

So What, Continued?

- FDA gave provisional 9-year approval, based on running a new study to establish clinical benefit during those 9 years
  - Who is going to enroll in the new study, and possibly be given a placebo, when a drug is available?
  - Suppose a new drug is proposed. Should new studies compare it to placebo, or to Aduhelm?
  - What endpoints should be used in a new study, given that previous study failed to show benefit for the cognitive test most commonly used?

Some Statistical Issues

- Original clinical outcomes were contradictory. Then the fishing started.
- Results had been unblinded already when the fishing started.
- Only patients in earliest stage of disease were tested, so no way to know if it works for those with longer term Alzheimer's.
- Future studies are likely to be biased because of who will be willing to enroll.

Part 3:
How statistics educators can help with data science ethics

Areas of ethical concern for statisticians**

- Ethics in data collection, quality and uses
- Ethical implementation of details in a study
- Issues of ethics during the analysis
- Ethics of reporting results
  - To clients
  - To the media and the public
- Teaching statistical literacy in all introductory statistics courses is an ethical obligation.

Facebook/Cornell Emotion Study**

- 2012 study, randomly selected 689,003 Facebook users, assigned to 4 groups.
- No informed consent!
- One group had negative news feed reduced; another had positive news feed reduced. Control groups had news feed randomly omitted. Study lasted one week.
- Use of negative and positive words used in subjects' own posts were measured.

Results from Cornell press release

"**News feed: Emotional contagion sweeps Facebook**"

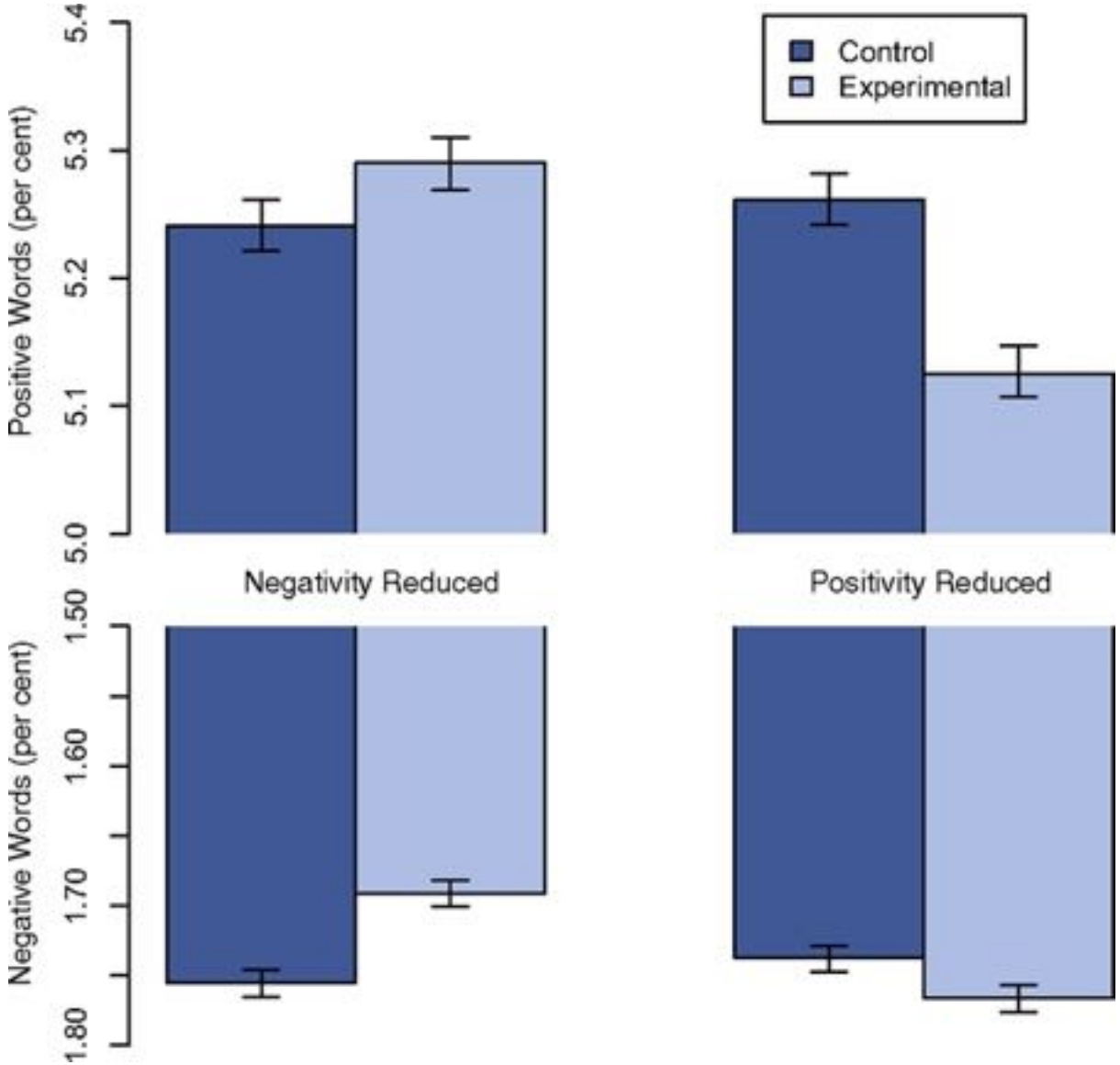
- "People who had positive content experimentally reduced on their Facebook news feed for one week used more negative words in their status."
- "When news feed negativity was reduced the opposite pattern occurred. Significantly more positive words were used in peoples' status updates."

https://news.cornell.edu/stories/2014/06/news-feed-emotional-contagion-sweeps-facebook

BUT, the actual results…

- Lowered positive posts:
  - Users' positive words decreased by 0.1% compared with control
  - $t(310{,}044) = 5.63$, $P < 0.001$, Cohen's $d = 0.02$
- Lowered negative posts:
  - Users' negative words decreased by 0.07%
  - $t(310{,}541) = 5.51$, $P < 0.001$, $d = 0.02$
- Remember, $n > 689{,}000$!

Misleading graphs….



Ethical Issues from this Study

- No informed consent
- Misleading graphs
- Confusion of statistical significance with practical significance (importance)
- Misleading reporting of results
- Authors justified small effect size as being of practical importance because of large population affected.

Ethics of Reporting Results

- Focus on magnitude, not p-values.
  - With big data, small effects have tiny p-values
- Include clear explanation of uncertainty.
- Don’t overstate the importance of results.
- Graphics should be clear, not misleading.
- Don’t imply causal connection if not justified.
- Media coverage should include all relevant results, not just most interesting or surprising.

Example: Reporting to client & media

- Suppose a client asks you to evaluate an online game for boosting children's math skills.
- Data provided include pre-post math and language scores, time spent studying each.
- Results: Math scores went up but language scores went down, and game was addictive.
- Are you ethically bound to report the negative consequences of using the game…
  - To the client?
  - In media requests?

Example: Hormone replacement therapy**

- Women's Health Initiative, randomized study comparing hormones with placebo.
- Surprising result was *increase* in risk of coronary heart disease in hormone group.
- Trial was stopped early, millions of women advised to stop taking HRT immediately.
- Large scale media attention on risks of heart disease and breast cancer from HRT.

But… Results from the original article

"Absolute excess risks per 10,000 person-years attributable to estrogen plus progestin were 7 more CHD [coronary heart disease] events, 8 more strokes, 8 more PEs [pulmonary embolism], 8 more invasive breast cancers, while absolute risk reductions per 10,000 person-years were 6 fewer colorectal cancers and 5 fewer hip fractures."

More results..

- Died of any cause during study:
  - Hormones: 231 of 8506 or 2.72%
  - Placebo: 218 of 8102 or 2.69%
- Adjusted for the time spent in the study, the death rate was slightly lower in the hormone group, with an annualized rate of 0.52% compared with 0.53% in the placebo group.

Ethical issue for reporting results

- The media and medical community focused on the surprising heart disease results
- In fact the hormone group fared *better* in many ways, including adjusted death rate.
- Were millions of women misled?
- If full results had been reported in the media, women could decide for themselves, for instance based on family or personal medical history.

Part 4:
Encouraging statistical literacy

Ethics in statistics education

- For training statisticians:
  - Include ethical considerations throughout their training
  - Include discussion of ethical issues as part of all data analysis projects, possibly dissertations as well*
- For educating all students:
  - Statistical literacy includes recognizing ethical issues
  - Emphasize topics students can use in their lives, to make informed decisions and recognize statistical errors

*Thanks to Eric Vance for this suggestion.

My Current Top 10 Important Literacy Topics

1. Observational studies, confounding, causation**
2. The problem of multiple testing**
3. Sample size and statistical significance**
4. Why many studies fail to replicate**
5. Does decreasing risk actually increase risk?
6. Personalized risk versus average risk
7. Poor intuition about probability and risk**
8. Using expected values to make decisions**
9. Surveys and polls – good and not so good
10. Confirmation bias

Poor intuition about probability & risk

- William James was first to suggest:
  - *intuitive* mind – quick, for survival
  - *analytical* mind – slow, thoughtful
- Psychologists have studied many ways in which we have poor intuition about risk and probability
- Recommended reading:
  - *Thinking, Fast and Slow* by Daniel Kahneman
  - Fast = intuitive mind
  - Slow = analytical mind

Confusion of the inverse: P(A|B) ≠ P(B|A)

- P(pos. test | disease) > P(disease | pos. test)
  especially for rare disease (doctors get this wrong)
- Prosecutor's fallacy in courtroom:

  P(innocent | evidence) vs

  P(evidence | innocent)
- COVID Example: UK, Delta variant (as of June 25)

  P(Double Vaccinated | Death) ≈ 0.43 = 50/117

  But P(Death | Double Vaccinated) is still very low!

If everyone was vaccinated, P(Vaccinated | Death) = 1.0!

Confusion of the inverse: DNA Example**

- Dan's DNA matches DNA at a crime scene. Only 1 in a million people have this specific DNA.
- There are 6 million people in the local area, so about 6 have this DNA.
- Is Dan almost surely guilty?

How to illustrate conditional probabilities

- P(DNA match | Dan is innocent)
  - ≈ 5 out of almost 6 million, extremely low!
  - *Prosecutor would emphasize this*

	Guilty	Innocent	Total
DNA match	1	5	6
No match	0	5,999,994	5,999,994
Total	1	5,999,999	6,000,000

DNA Example, continued

- But… P(Dan is innocent | DNA match)

≈ 5 out of 6, fairly high!

- *Defense lawyer would emphasize this*

	Guilty	Innocent	Total
DNA match	1	5	6
No match	0	5,999,994	5,999,994
Total	1	5,999,999	6,000,000

More general version

	Guilty	Innocent	Total
Match rare evidence	1	5	6
No match	0	5,999,994	5,999,994
Total	1	5,999,999	6,000,000

- P(Innocent | evidence match) = 5/6
- P(Evidence match | innocent) = 5/5,999,999

Prosecutor's Fallacy

- P(DNA match | innocent) = 5/5,999,999 *very low*

  P(match evidence | innocent)

  Prosecutor would emphasize this

- P(innocent | DNA match) = 5/6 *high*

  P(innocent | match evidence)

  Defense lawyer would emphasize this

- Jury needs to understand this difference!

Suggestions for Statistics Educators

- Train all students to think about ethics:
  - Propose and discuss ethical issues in class
  - Include ethics section in all data analysis assignments
  - Ask *why* before *how*
- For non-statistics students, teach literacy issues such as:
  - Recognizing multiple analyses and selective reporting
  - Trade off in risks
  - Trade offs in society benefits versus personal rights
  - Confusion of the inverse

Preparing Statistics Majors for Jobs

- It's not true that numbers are just numbers – learning subject matter knowledge is crucial!
  - Rebecca Nugent: "If you are not talking to content experts early and often, you will be starting over."
- Often the statistician will be the most objective member of a multi-disciplinary team.
- Need to speak up as a member of a team.
- Ask good questions! (Thanks, Allan Rossman)
  - Including "why" before "how"

Summary and Conclusions

- Statistics educators should play a major role in data science ethics.
- Statistics majors at all levels should be thinking about ethics along with technical issues.
- Students who take one course in statistics should be able to identify bias, confounding, selective reporting, and other misleading issues in statistical studies.
- Statistics educators can change the world for the better!

THANK YOU

Contact info:

jutts@uci.edu

http://www.ics.uci.edu/~jutts

UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE