

STATISTICS IN THE WORLD: ESSAY 3
Jo Hardin Fall 2021

The third paper will be a research paper describing how statistics (as a discipline, as a set of data analysis tools, as individual people, etc.) are/were involved in creating or reinforcing systemic inequality or social injustice. 1500 - 2000 words (concise is best!).¹

INSTRUCTIONS

Your final paper will center around a research question (that will become a thesis statement) which will be argued using sources that you find. The research question will continue to evolve over the life of the project, but you should keep coming back to the (evolving) idea you want to argue. In addition to a **succinct and arguable thesis statement**, your paper should include:

1. Thesis statement somewhere early on and clear to your reader.
2. Details of the event of interest.
3. Historical context (*why* the thing you describe happened / was able to happen).
4. (Optional) You might also give the reader an idea of possible solutions (how to make it better) or maybe a counter argument.
5. A potential audience for your paper is a STEM undergraduate who feels disempowered by the traditional canon. If your paper speaks to a different audience, please talk to me before turning in your assignments.

IDEAS FOR GETTING STARTED INCLUDE:

- Some of the founders of the discipline of statistics were eugenicists. (Galton and the origins of correlation, fame and genius within families; Karl Pearson and goodness-of-fit tests, differences among populations; RA Fisher “The Genetical Theory of Natural Selection” 1930)
- Original medical studies were done entirely on white men.
- 23 & me has information primarily on white, thin, high income Americans. <https://podcasts.apple.com/us/podcast/freakonomics-radio/id354668519?i=1000438218337>
- Contributions to the field of statistics that were by individuals in marginalized roles were attributed to white men.
- Corollary: because science is dominated by white men, data exist primarily on their contributions. Thus, contributions by non-white men get overlooked. Recently (10/2/2018), Donna Strickland was awarded the Nobel Prize in Physics. Just a few months prior (5/23/2018) Wikipedia refused to allow her page to be created. As reported in *The Atlantic*, “ ‘This submission’s references do not show that the subject qualifies for a Wikipedia article.’ Strickland, it was determined, had not received enough dedicated coverage elsewhere on the internet to

¹Thanks to Vin De Silva, Jorge Moreno, Janice Hudgings, Tomas Summers Sandoval, Pam Bromley, and Kara Wittman for their ideas and suggestions on this essay prompt.

warrant a page.” <https://www.theatlantic.com/science/archive/2018/10/nobel-prize-physics-donna-strickland-gerard-mourou-arthur-ashkin/571909/> [The problem here is about human influence into data and algorithms.]

- The larger issue: current algorithms and AI protocols reinforce existing structures and are often biased against marginalized communities. (Summary and ideas here: <https://www.infoq.com/presentations/unconscious-bias-machine-learning>.)
- Statistical models were used as part of redlining.
- Tuskegee Study of Untreated Syphilis in the Negro Male or Willowbrook Study.
- Collider bias, e.g., see “Why Statistics Don’t Capture the Full Extent of the Systemic Bias in Policing” by Laura Bronner, <https://fivethirtyeight.com/features/why-statistics-dont-capture-the-full-extent-of-the-systemic-bias-in-policing/>
- Using “statistics” to show connections between race and disease sets up a structure of biological differences instead of focusing on systemic disparities in healthcare and other socio-economic factors. See “Racial Health Disparities and Covid-19 – Caution and Context” in **New England Journal of Medicine**, <https://www.nejm.org/doi/full/10.1056/NEJMp2012910>.
- Statistical models used to argue biological determinism (e.g., see **The Mismeasure of Man** by Stephen Jay Gould and related controversy ... <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001071> and http://blogs.nature.com/news/2011/06/did_stephen_jay_gould_fudge_hi.html)
- If scientific knowledge, data-based arguments, and logical & critical reasoning are *sound* ways of discovery, why were some of the statistics (STEM generally) founders so prone to racist / misogynist / ableist / etc. ideas? That is, what can we *know* from science and what can we *never know* from science?
- Weapons of Math Destruction (Cathy O’Neil) are 1. opaque, 2. at a large scale, 3. destructive. What WMDs are out there? Wellness programs, employment tests, college applications, credit scores, insurance, etc.

ASSIGNMENT TIMELINE

Meeting with Writing Partner. At some point in the process, you must meet with Chris Meng about the research paper. You may meet with him before you start, after you've written your micro-essay, or after your first draft of the paper is complete. He will provide times for meeting throughout the last few weeks of the semester, but it is your responsibility to sign up to meet with him at some point during your writing process.

Micro-essay. Write a proposal for your research and paper. Your proposal should include the specific research question of interest with possible responses (provide three different possible responses indicating different directions your research might go). Additionally, you should provide some ideas (or possibly sources) for the evidence you will use to address the problem. 500 words maximum.

Thursday, November 11, midnight to Sakai.

Annotated Bibliography. Submit a bibliography of three to five scholarly articles that describe either the event of focus or the impact of the event you describe. Identify each article in one of the standard bibliography formats, and write a **few sentences** for each article summarizing its content (no need to give a detailed summary about the entire work). Additionally, provide a few sentences about what each source says about your research question.

Sunday, November 21, midnight to Sakai. Note: Thanksgiving weekend!!!

Version 1. This is a complete and polished essay. Your thesis statement should be clear, but it may have evolved through the process. Prof Hardin will read the Version 1 essays you submit.

Sunday, November 28, midnight to Sakai.

Peer review. Groups of three. Same format as before, but you will meet outside of class.

Monday, December 3 - Wednesday, December 5.

Class presentation. Five minutes per person, everyone will present their work to the class.

Tuesday, December 7.

Version 2. This is a complete and polished essay. The essay should answer your research question in the form of an argument.

Friday, December 10, midnight to Sakai.

OR

Poster. Instead of a final paper (Version 2), you are invited to turn your project into a poster as part of the STEM Equity Symposium.

Wednesday, December 8, afternoon.

ANNOYING FORMATTING DETAILS

Really annoying requirements that will again make it easier to read many papers:

- Identify your overall thesis (the core argument around which the entire essay is organized) **and bold-face it**.
- Identify the topic sentence of each paragraph and underline or *italicize* it.
- Name the file you upload to Sakai like this: Paper#.LastName.doc, as in Paper2.Gallup.doc.
- Be sure that your name, title of the paper, and date are on the manuscript itself.
- Double-spaced with 12 point type and at least 1 inch margins to write comments in.
- Any reference / citation formatting is fine, but please be consistent.
- Print a word-count at the end of your paper.

AN ASIDE

It is valuable to point out that statistics / statisticians have also been forces of good in myriad ways.²

- Pearson and Fisher were eugenicists, but their work in statistics and genetics played pivotal roles that have led to the modern, scientific, debunking of eugenics.
- Cyril Burt both made contributions to statistics and misused statistics to lie about his “data” in his studies of heredity and IQ in twins. However, statisticians and their professional societies have explicit policies against inventing and modifying data.
- Statistical models were used for redlining... and also in court cases against it. One could investigate the statistical arguments brought by Thurgood Marshall and the NAACP to argue against segregation and “separate but equal”. Or the current statistically-based arguments against race-based gerrymandering (that recently won a court case in GA).

There have clearly been misuses of statistics in support of inequality, but we should be careful not to assume that the entire field of statistics is designed to produce systemic inequality.

²Thanks to Paul Velleman for pointing out contributions of statistics to producing systemic *equalities*.