

An example of combining data analysis with ethics-oriented reflection

Julie Vu
06 August 2024



HARVARD
Faculty of Arts and Sciences
DEPARTMENT OF STATISTICS

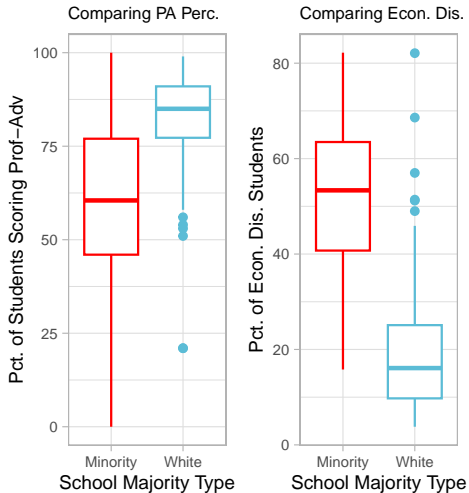
IMPACT OF STATISTICAL PRACTICE

- ASA's Ethical Guidelines for Statistical Practice are directed to **all** those who engage in statistical practice.
- With the growing impact of statistics and data science on society, an early focus on ethical principles becomes ever more critical.
- How can we substantively incorporate data science ethics into an introductory course?
 - ◊ Explicitly draw connections between statistical concepts and the associated ethical dimensions.
 - ◊ Incorporate specific examples of how statistical practice impacts society.

THE DATA

- Research has shown that variation in standardized test scores is often explained by factors such as race and economic class.
- Let's investigate whether data from the 2018 Mathematics MCAS test indicate evidence of poverty-based and/or race-based achievement gaps.
- Key variables:
 - ◊ Percentage of students scoring at the Proficient or Advanced levels
 - ◊ Percentage of students from an economically disadvantaged background
 - ◊ Whether a school is majority-minority or majority-white

THE DATA ANALYSIS



Estimate the association between PA_perc and minority when adjusting for potential confounders.

- Model without econ_dis:

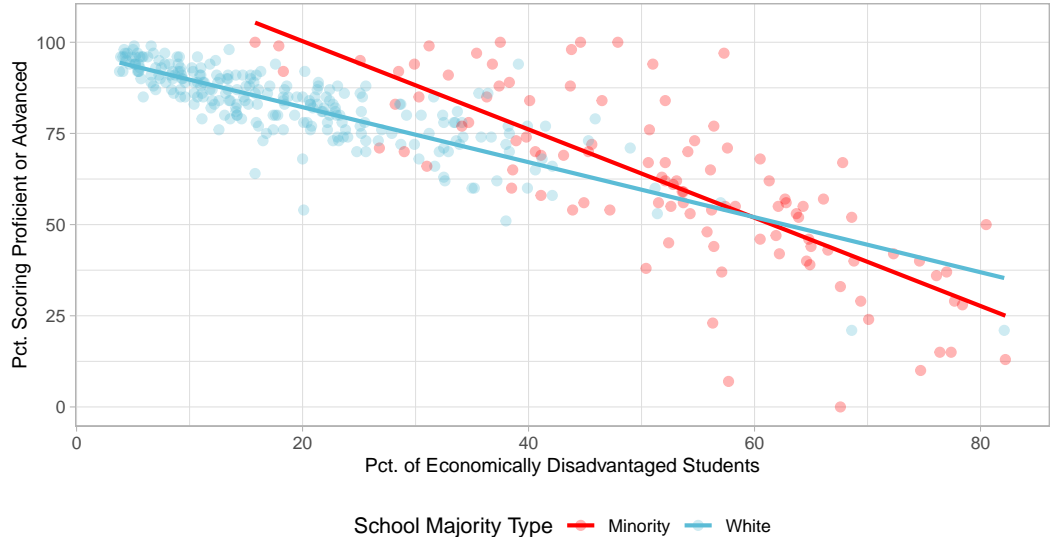
term	estimate
<chr>	<dbl>
intercept	-100.081
majority: White	14.974
average_math_class_size	1.350
attendance_rate	1.523
student_teacher_ratio	0.133

- Model including econ_dis:

term	estimate
<chr>	<dbl>
intercept	31.598
majority: White	-5.892
econ_dis	-0.748
average_math_class_size	0.709
attendance_rate	0.643
student_teacher_ratio	-0.091

THE DATA ANALYSIS...

Comparing the Association between Test Performance and Economic Disadvantage by School Type



CHALLENGES FOR STUDENTS

Prompt.

Suppose these results will be discussed in a future meeting of the Racial Imbalance Advisory Council. . . which advises on matters related to providing access to effective educational programs for all students in the state regardless of race or socioeconomic class.

Prepare a statement, no more than ten sentences long, summarizing the main findings. . . Be sure to use language that is **accessible to a general audience** and make specific references to previous numerical results.

- Reading for detail: data background, variable descriptions
- The technical material: understanding regression models (coefficient interpretation, interaction)
- Drawing meaning from the analyses
- Constructing an effective narrative

ETHICS-ORIENTED REFLECTION

"Statisticians should contemplate, and be sensitive to, the manner in which information is framed to avoid disproportionate harm to vulnerable groups."

Students read "Racial Health Disparities and Covid-19—Caution and Context" by Chowkwanyun and Reed Jr. (2020).

- Data in a vacuum may support explanations grounded in racial stereotypes
- Perception of certain social problems as "racial" has been used to rationalize neglect and funding cuts

"Why is caution necessary when presenting findings from the MCAS data?"

- May be used to argue for inherent differences in intelligence/ability

"How might you contextualize the findings when discussing them with a general audience?"

- Discuss important confounders such as socioeconomic status, disparities in quality of education, . . .
- Connect analysis to research on racial bias in standardized testing

STRENGTHS OF THE MCAS EXAMPLE

- Approachable (and engaging) data context
 - ◇ Students are familiar with standardized tests
 - ◇ Students are interested in data connected to social issues
- Rich dataset that lends itself well to visualization, inference, regression
- Strong connection to data ethics and data justice
 - ◇ Students are not typically exposed to thinking about what happens after an analysis is completed
 - ◇ Direct example of a situation where inappropriate contextualization would be irresponsible
 - ◇ Can particularly resonate with students who are part of underrepresented groups and/or who come from economically disadvantaged backgrounds

ACCESSING MATERIALS

The MCAS dataset and talk materials are available on GitHub:

https://github.com/JulieVu/mcas_jsm_2024/

