Emily Hsiao

ehsiao@utexas.edu

Education

University of Texas, Austin
Ph.D. Student, Statistics — 2021 - Expected May 2026
Advisor: Dr. Layla Parast
Committee: Dr. Layla Parast, Dr. Roger Peng, Dr. Jay Bartroff, Dr. Beth Ann Griffin

University of California, Berkeley
B.A. with Honors, Data Science — 2017 - 2021
Thesis: Auditing Search Engine Bias - Google and Duckduckgo

Research Experience

Department of Statistics, UT Austin
Graduate Researcher — Fall 2023 - current
Advisor: Layla Parast
Developing a general framework to test assumptions required in current surrogate marker validation techniques to avoid the surrogate paradox using nonparametric statistic methods.

Los Alamos National Lab
Graduate Student Researcher — Summer 2023
Advisor: John Tipton
Developed a model for forecasting mosquito abundance by accounting for sampling effort using temporal change of support. Used for an epidemiological model for predicting cases of mosquito-borne diseases such as denggue and West Nile Virus.

Department of Statistics, UT Austin
Graduate Researcher, Calder Research Group — Summer 2022 - Spring 2023
Under Dr. Kate Calder, used spatial point process models to model crime in Columbus, Ohio with spatially smoothed and areal neighborhood characteristics as model inputs.

Teaching Experience

Teaching Assistant, UT Austin, Department of Statistics and Data Science

SDS 313: Introduction to Data Science	Fall 2023
SDS 302F: Foundations of Data Analysis	Fall 2021, Fall 2022, Spring 2023
SDS 321: Introduction to Probability and Statistics	Spring 2022
SDS 384: Design Principles and Causal Inference	Fall 2024

Volunteer Instructor, Texas Prison Education Initiative

Math 305G: Precalculus	Fall 2023, Fall 2024
College Prep Math	Spring 2022, Fall 2022

Assistant Instructor, UT Austin, Department of Statistics and Data Sciences

SDS 320E: Elements of Statistics	Spring 2024

Teaching Assistant, UC Berkeley, Department of Statistics

Stat 88: Probability and Mathematical Statistics for Data Science	Fall 2020, Spring 2021
Stat 140: Probability for Data Science	Spring 2020

Industry Experience

Facebook
Software Engineering Intern — Summer 2020
Infra/Backend software engineering intern. Coded in python and thrift writing a service implementation of data mining algorithms.

Publications

Hsiao E, Tian L, Parast L. Avoiding the Surrogate Paradox: An Empirical Framework for Assessing Assumptions. *Journal of Nonparametric Statistics,* In Press.

Calder, Pinchak, Browning, Carter, **Hsiao**, Boettner, Tarrence, Bellair. Mobility Network-Based Measurement of Local Collective Efficacy and its Consequences for the Spatial Patterning of Violent Crime *Journal of Quantitative Criminology,* Accepted.

Hsiao E, Tian L, Parast L. Resilience Measures for the Surrogate Paradox. Submitted.

Awards

Outstanding Data Science Undergraduate Award
UC Berkeley, Department of Data Science — May 2021

Invited Presentations

Rank-Based Identification of Surrogates in Small Ebola Studies Symposium — June 2024
Hangzhou International Conference on Frontiers of Data Science — July 2024

Conference Presentations

ENAR International Biometric Society Meetings, New Orleans, LA — March 2025

Journal Referee

Health Services and Outcomes Research Methodology
Journal of the Royal Statistical Society: Series B

Community Outreach

Opportunity Through Data
Data Science Class Teacher — Spring 2021
Taught classes in data science techniques (python, pandas, regression, etc) to underrepresented and under-resourced high school students.

Bridging Berkeley
Middle School Student Mentor — Spring 2018
Mentored middle school students, especially students who will be first-generation college students, in math and career planning at a youth center.

Technical Skills

R, LaTeX, Java, C, Python, Git

Professional Memberships

Institute of Mathematical Statistics
ENAR of the International Biometric Society
American Statistical Association

SOFTWARE `SurrogateParadoxTest`, an R package nonparametrically assess assumptions necessary to prevent the surrogate paradox through hypothesis tests of stochastic dominance, monotonicity of regression functions, and non-negative residual treatment effects