

Emily Hsiao
ehsiao@utexas.edu

EDUCATION	University of Texas, Austin Ph.D. Student, Statistics Advisor: Dr. Layla Parast Committee: Dr. Layla Parast, Dr. Roger Peng, Dr. Jay Bartroff, Dr. Beth Ann Griffin	2021 - Expected May 2026
	University of California, Berkeley B.A. with Honors, Data Science Thesis: Auditing Search Engine Bias - Google and Duckduckgo	2017 - 2021
RESEARCH EXPERIENCE	Department of Statistics, UT Austin Graduate Researcher <i>Advisor: Layla Parast</i> Developing a general framework to test assumptions required in current surrogate marker validation techniques to avoid the surrogate paradox using nonparametric statistic methods.	Fall 2023 - current
	Los Alamos National Lab Graduate Student Researcher <i>Advisor: John Tipton</i> 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 dengue and West Nile Virus.	Summer 2023
	Department of Statistics, UT Austin Graduate Researcher, Calder Research Group 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.	Summer 2022 - Spring 2023
	Teaching Assistant , UT Austin, Department of Statistics and Data Science SDS 313: Introduction to Data Science SDS 302F: Foundations of Data Analysis SDS 321: Introduction to Probability and Statistics SDS 384: Design Principles and Causal Inference	Fall 2023 Fall 2021, Fall 2022, Spring 2023 Spring 2022 Fall 2024
TEACHING EXPERIENCE	Volunteer Instructor , Texas Prison Education Initiative Math 305G: Precalculus College Prep Math	Fall 2023, Fall 2024 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 Stat 140: Probability for Data Science	Fall 2020, Spring 2021 Spring 2020

INDUSTRY	Facebook	
EXPERIENCE	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. <i>Journal of Nonparametric Statistics</i> , In Press.	
	Calder, Pinchak, Browning, Carter, Hsiao , Boettner, Tarrence, Bellair. Mobility Network-Based Measurement of Local Collective Efficacy and its Consequences for the Spatial Pattern-ing of Violent Crime <i>Journal of Quantitative Criminology</i> , In Press.	
	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	Rank-Based Identification of Surrogates in Small Ebola Studies Symposium	June 2024
PRESENTATIONS	Hangzhou International Conference on Frontiers of Data Science	July 2024
CONFERENCE	ENAR International Biometric Society Meetings, New Orleans, LA	March 2025
PRESENTATIONS		
JOURNAL	Health Services and Outcomes Research Methodology	
REFEREE	Journal of the Royal Statistical Society: Series B	
COMMUNITY	Opportunity Through Data	
OUTREACH	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 stu-dents, in math and career planning at a youth center.	
TECHNICAL	R, L ^A T _E X, Java, C, Python, Git	
SKILLS		
PROFESSIONAL	Institute of Mathematical Statistics	
MEMBERSHIPS	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