

Sean Mulherin

Research Interest

Statistics, Time Series, Point Processes, Finance, Education

Education

Ph.D. Statistics, UCLA	3/2025 - Present
Advisor: Prof. Frederic Schoenberg	
M.S. Applied Statistics and Data Science, UCLA	9/2023 - 3/2025
<i>Testing for Causal Clustering in K-12 Student Discipline</i>	
Advisor: Prof. Frederic Schoenberg	
M.A.T. Secondary Mathematics, UNC - Chapel Hill	5/2019 - 9/2020
Advisor: Dr. Josh Corbat	
B.S. Mathematics, North Carolina State University	9/2015 - 5/2019

Professional Experience

UCLA Department of Statistics and Data Science	9/2024 - Present
○ Teaching Assistant	
• Introduction to Statistical Reasoning	
• Societal Impacts of Data	
• Data Technologies for Data Scientist	
○ Student Ambassador	
SoSystems	6/2025 - Present
○ Lead AI Architect	
• Designed and led a production-grade, multi-stage AI pipeline that applies statistical reasoning, semantic similarity analysis, and structured inference to transform unstructured government regulations into validated, machine-readable compliance outputs.	
• Incorporated confidence scoring, clustering, deduplication, and validation checks to manage uncertainty and ensure reproducible, trustworthy results in a high-stakes regulatory setting.	

Brentwood School	8/2024 - 8/2025
◦ Physics Faculty	
◦ Cross Country Coach	
◦ Academic Advisor	
Jackson Hole High School	8/2021 - 8/2023
◦ Math Faculty: Geometry, AP Prep Algebra II, Trigonometry/Precalculus	
◦ Cross Country & Track Coach	
◦ Chess Club Coach	
Mountain Academy of Teton Science Schools	8/2020 - 8/2021
◦ Lead Math Faculty	
• Algebra	
• Geometry	
• IB Applications & Interpretations	
◦ Academic Advisor	
Carrboro High School	8/2019 - 8/2020
◦ Student-Teacher	
• Geometry	
• AP Calculus AB	
• AP Calculus BC	
North Carolina State University Tutorial Center	10/2016 - 5/2018
◦ Math Tutor: Calculus I-III, Foundations of Advanced Mathematics, Differential Equations I, Mathematical Probability	

Research Experience

Current Research Project *In-Progress*
Accurate Estimation of Cross-Productivity in Spatial-Temporal Hawkes Models of Infectious Disease Using Covariates

Conventional maximum likelihood estimates of cross-productivities in spatial-temporal Hawkes processes are known to be substantially biased. Here, two methods are investigated, one involving the use of human mobility data, and the other using geographical distance. In both methods, Hawkes models are fit separately to each location, yielding local parameters. Next, keeping local parameters fixed, cross-productivities are estimated via penalized negative binomial likelihood maximization with elastic net regularization. The methods are tested on simulations and applied to observed Covid-19 data in the United States during 2020. The resulting parameter estimates are shown to improve model fit by reducing spurious connections across locations and dormant mobility pipelines, with distance and mobility penalization

methods offering superior fits relative to the unpenalized baseline method, outperforming the baseline by 1.86% and 2.25%, using root mean squared error as a metric.

UCLA Master's Thesis, Testing for Causal Clustering in K-12 Student Discipline 2024

The degree to which causal contagion explains the event of student misconduct is investigated. A test introduced by Kresin (2023) and McGovern (2024) is applied, wherein likelihood-ratio tests are performed using information gain statistics to compare the fit of a Neyman-Scott model to that of a Hawkes model. Data records the number of disciplinary actions for K-12 students grouped by day, school, and school year spanning from 2016 to 2023. Evidence of causal clustering appears to vary across grade levels and school years. Out of the twelve different school/year combinations tested, seven exhibited statistically significant evidence of causal triggering. Interestingly, both frequency and magnitude were found to hold strong governance over the conclusion of hypothesis tests.

Advanced Studies Institute in Mathematics of Data Science & Machine Learning 2024

Sponsored by the National Science Foundation, I was awarded a research grant for travel to Uzbekistan to participate in a specialized research workshop focusing on the mathematics of machine learning. Topics covered include model-based clustering, Hawkes point processes, benign overfitting, generalization, double descent, and mirror descent.

University of North Carolina, Chapel Hill Research Project 2020

Assessing Collaboration and Critical Thinking Opportunities in Online Learning at the Secondary Level during the COVID-19 pandemic

This paper studies the instructional strategies and tools secondary school that teachers used during the COVID-19 quarantine to promote collaboration and critical thinking amongst their students through virtual learning. A sample of the literature about online instructional techniques and critical thinking or collaboration is reviewed. We collected data from 36 secondary instructors via online surveys and virtual interviews. Results show a strong positive correlation between teachers who report higher amounts of participation and the assigning of group work.

North Carolina State University, College of Design 2017 - 2018

As a research assistant, I collected data pertaining to the efficacy of healthy diets on the social, emotional, and academic performance of elementary school students.

Portfolio Projects

Financial Modeling 2024

Programmed a web application that provides two tools operating at the confluence of statistics and finance. The first is a forecasting tool that uses the Prophet model to fit and predict the daily adjusted closing price of a user-specified asset. The second tool is one for portfolio optimization and valuation wherein the user inputs multiple assets and relevant performance metrics are displayed such as volatility, return, weights, and so on.

Forecasting Selling Price of Houses in the U.S. 2024

Programmed an interface for users to input a city and forecast period to observe and explore the forecasted trends and prediction metrics computed by the Prophet model. This model leverages concepts from SARIMAX models to accurately fit and predict time series data. Moreover, housing prices are atypically volatile, so this model does particularly well at producing accurate home price forecasts.

An Artificial Neural Network Approach to Identifying Diabetes Risk Status 2023

Programmed an artificial neural network from scratch to classify a user's risk of developing type II diabetes after completing a 21 question survey. The model was trained using CDC data and achieves 84% accuracy in its validation-set predictions.

Tracking Global Carbonization 2023

Conducted a comprehensive statistical analysis of the current state of global carbon dioxide emissions. Data was obtained from the United National Development Program and analyzed using R. Advanced data visualizations were displayed to portray interpretable results.

A Classification Analysis on Breast Cancer Tumors 2023

Evaluated various models that were built to classify breast cancer tumors as malignant or benign. Models compared include: linear discriminant analysis, quadratic discriminant analysis, support vector machines, logistic regression, random forests, Naive Bayes, and KNN. The most optimal model used linear discriminant analysis to predict with 97% accuracy on the validation set.

Talks & Presentations

International Conference on Infectious Disease Dynamics - Poster Presentation 2025

DataFest Conference Guest Speaker - Introduction to R 2024

DataFest Conference Guest Speaker - Data Cleaning and Wrangling in R 2024

Appointments & Awards

UCLA Statistics Outstanding Masters Student Award 2025

UCLA Math and Physical Sciences Council Member 2023 - 2025

National Institute of Statistical Sciences GSN Council Member 2023 - 2025

NCAA Division I Cross Country & Track Athlete 2015 - 2019

Note: all of my data science projects, academic papers, and lecture notes from my teachings can be found on my online portfolio linked in the header. All projects are programmed using Python, R, and HTML with Github as the cloud platform.