Sean Mulherin

Los Angeles, CA
smulherin519@gmail.com
seanmulherin.github.io
336-681-0821

Research Interest

Statistics, Time Series Analysis, Point Process Analysis, Education, Social Statistics

Education

M.S. Applied Statistics and Data Science, *UCLA*   2023 - Present

    Testing for Causal Clustering in K-12 Student Discipline

    Advisor: Dr. Frederic Schoenberg

M.A.T. Secondary Mathematics, *University of North Carolina, Chapel Hill*   2020

    Advisor: Dr. Josh Corbat

B.S. Mathematics, *North Carolina State University*   2019

Teaching Experience

Brentwood School   2024 - Present
- Physics Faculty
- Data Science Expository Course
- Cross Country Coach
- Academic Advisor

UCLA Department of Statistics and Data Science   2024 - Present
- Reader/Grader
- Student Ambassador

Jackson Hole High School   2021 - 2023
- Math Faculty: Geometry, AP Prep Algebra II, Trigonometry/Precalculus
- Cross Country & Track Coach
- Chess Club Coach

Mountain Academy of Teton Science Schools   2020 - 2021
- Lead Math Faculty: Algebra, Geometry, IB Applications & Interpretations
- Health & Wellness Teacher
- Academic Advisor

Carrboro High School   2019 - 2020
- Student-Teacher: Geometry, AP Calculus AB, AP Calculus BC

North Carolina State University Tutorial Center   2016 - 2018
- Math Tutor: Calculus I-III, Foundations of Advanced Mathematics,
    Differential Equations I, Mathematical Probability

Research Experience

Current Research Project   *In-Progress*
Mitigating False Cross-Excitation in Multivariate Hawkes Processes Models of Infectious Diseases

This paper sets out to investigate a method to mitigate the occurrence of false cross-excitation of infectious disease transmission in multivariate Hawkes process (MHP) models. Conventional Hawkes models neglect geographic distance, and therefore are unable to penalize the likelihood of cross-excitation where contagion is improbable due to geographic barriers. The proposed method uses maximum-likelihood estimation to fit independent MHP models to each location, yielding intracity productivity rates. Keeping these intracity models fixed, an overall MHP model is fitted to estimate intercity cross-productivity rates. COVID-19 data are used to model intracity productivity, while mobility data are used to fit intercity productivity. The optimized productivity matrix reflects penalized cross-excitations between dormant city-pairs and is shown to improve the model fit by reducing spurious connections across distant locations.

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 traveled to Uzbekistan to participate in a two-week 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.

An Ethereum Regression Analysis 2023

Built and compared various regression models to predict the USD-ETH price in the year 2030. Models include: linear, quadratic, cubic, exponential, and logarithmic regression.

Appointments

National Institute of Statistical Sciences GSN Council Member	2023 - Present
UCLA Statistics Graduate Student Association, VP of External Affairs	2023 - Present
UCLA Math and Physical Sciences Council Member	2023 - Present
DataFest Conference Guest Speaker - Introduction to R	2024
DataFest Conference Guest Speaker - Data Cleaning and Wrangling in R	2024
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.