William (Zehao) Qian

401-585-6820 zehao_qian@brown.edu william-z-qian dukechain2333 portfolio.williamq.me

Education

Brown University **Sep. 2023 – May. 2025**
Sc.M. in Biostatistics (Thesis) *Providence, Rhode Island*

- GPA 3.83/4.00.

Shanghai Lixin University of Accounting and Finance **Sep. 2019 – Jun. 2023**
B.S. in Data Science and Big Data Technology *Shanghai, China*

- GPA 3.87/4.00.
- Awarded for National Scholarship.
- Graduated with distinction.

Publications

- J. Chen, Z. Hu and **Z. Qian**, "Research on Malicious URL Detection Based on Random Forest," 2022 14th International Conference on Computer Research and Development (ICCRD), Shenzhen, China, 2022, pp. 30-36, doi: 10.1109/ICCRD54409.2022.9730451.

Relevant Coursework

- Causal Inference & Missing Data
- Bayesian Statistical Methods
- Statistical Inference
- Applied Generalized Linear Models
- Longitudinal Data Analysis
- Multilevel Data Analysis
- Statistical Learning/Big Data
- Practical Data Analysis
- Statistical Programming with R
- Data Structure and Algorithms
- Optimizing Theory and Algorithms
- Multivariate Statistical Analysis

Research & Experience

Research Assistant **Jun. 2024 – Current**
Brown University & Boston Medical Center; Prof. Stavroula Chrysanthopoulou *Providence, Rhode Island*
HEAL Data2Action (D2A) Program - Emulator of RESPOND Model

- Utilized GEE, GLME, MERF and LSTM models to emulate the outcome of RESPOND model;
- Advanced data transformation and model tuning techniques such as Hyperband search are used to conduct comparisons between models to find out the one with the best performance;
- Constructed a model fitting and predicting pipeline for Boston Medical Center's research workflow.

Research Assistant **Mar. 2024 – Current**
Brown University; Prof. Alice Paul *Providence, Rhode Island*
A Causal Approach to Fair Predictive Modeling via Penalized Maximum Likelihood Estimation

- Utilized path-specific effects (PSEs) and penalized maximum likelihood estimation (PMLE) to address discrimination in datasets, ensuring fairer predictive outcomes;
- Employed the g-formula to estimate causal parameters, such as natural direct and indirect effects, for decomposing the total effect of variables on outcomes;
- Generalized this algorithm and developed a python package.*(Link)*

Teaching Assistant **Mar. 2024 – May. 2024**
Brown University; Applied Regression Analysis; Prof. Anarina Murillo *Providence, Rhode Island*

- Developed student understanding of statistical concepts by leading lab sessions and facilitating practice with R;
- Provided individualized support during office hours, guiding students on data analysis, research design, and statistical methodology selection;
- Evaluated student comprehension of statistical methods through meticulous grading of homework and exams.

Junior Data Analyst Intern **Jul. 2022 – Nov. 2022**
PricewaterhouseCoopers Shanghai Acceleration Center *Shanghai, China*

- Designed and implemented the database migration scripts and added them to the project pipeline, improving the process deployment efficiency by 50%;
- Optimized SQL queries for 60 million rows of data, reducing their execution time to one-third of the original;
- Analyzed the features and distribution of user data, providing insights from a data science perspective;
- Advanced the project launching date by one day by fixing bugs from a business perspective.

Research Assistant **Jun. 2021 – Sep. 2021**

University of Chicago; Prof. Nick Feamster *Remote*

Research on Malicious URL Detection Based on Random Forest

- Cleaned the experiment data to remove illegal values and split it into training and testing datasets;
- Trained the Random Forest model, and compare it with baseline built on machine learning algorithms such as SVM, Decision Tree, and XGBoost;
- Pruned the trained Random Forest Model to obtain the optimistic classification accuracy within the minimum model size.

Research Assistant **Nov. 2019 – Jan. 2023**

Enligence Laboratory of Shanghai Lixin University of Accounting and Finance; Prof. Gaoyu Zhang *Shanghai, China*

- Conducted experiments under the instruction of laboratory tutor;
- Participated in various professional skills and academic competitions and gained multiple prizes.

Course Projects

Leukemia Cancer Treatment Survival Analysis from a Bayesian Perspective

Bayesian Survival Analysis, MCMC, Cox Model, Stan *Course: Bayesian Statistical Methods*

- Applied an adjusted Bayesian Cox model with hierarchical priors to improve the flexibility and reliability of survival analysis for leukemia cancer treatment data;
- Utilized Markov Chain Monte Carlo (MCMC) algorithms in Stan to draw posterior samples and estimate parameters of the Bayesian Cox model, ensuring accurate parameter estimation and model fit;
- Compared survival probabilities between treatment and placebo groups, demonstrating the effectiveness of the new treatment in improving patient survival times.

A Complete Introduction to ResNet

Deep Learning, ResNet, Model Optimization *Course: Statistical Learning/Big Data*

- Provided a comprehensive overview of the basic theory behind Residual Networks (ResNet), including its structure and significance in deep learning;
- Discussed advanced applications and variations of ResNet, offering insights into its practical uses and potential improvements;
- Presented experimental results and analysis, showcasing the performance and advantages of ResNet over traditional neural networks.

Opioid Overdose Problems in the United States: Insights from Prescribing & Overdose Death Rates

Data Mining, Data Visualization, Practical Data Analysis *Course: Statistical Programming with R*

- Conducted a comprehensive analysis of Medicaid opioid prescribing rates, identifying significant variations in opioid prescriptions across different geographic regions and plan types;
- Performed an in-depth study of data from opioid treatment program providers, focusing on the availability and distribution of treatment resources;
- Analyzed the provisional drug overdose death counts to uncover emerging trends and patterns in opioid-related fatalities, highlighting key areas of public health intervention.

A Comparative Analysis of Interval Estimation Techniques in Small Sample Research

Wald Interval, Agresti-Coull Interval, Simulation *Course: Statistical Learning/Big Data*

- Conducted a comprehensive statistical analysis, focusing on Wald and Agresti-Coull Intervals, to evaluate their efficacy in various scenarios;
- Utilized simulations to compare interval performances against a 95% nominal coverage, emphasizing practical applications in clinical trials and market research;
- Given a unique interpretation from the Bayesian perspective.

Extracurricular

Master's APC Student Representative **Sep. 2024 – May. 2025**

Brown University, School of Public Health, Department of Biostatistics *Providence, Rhode Island*

- Solicit concerns from students to be presented to the Committee;
- Get student feedback on current issues being discussed in the Committee;
- Facilitate the communication between the committee and the students as needed.