Kaizhong Mu

Phone:5303648880 | Email: Kaizhong_mu@brown.edu | Address: 1st, Providence, RI

EDUCATION

Brown University   Providence, RI
Master of Science in Biostatistics ⎪ GPA: 4.00/4.00   09/24 – Present
University of California, Davis   Davis, CA
Bachelor of Science in Statistics and Economics ⎪ CGPA: 3.73/4.00 ⎪ Dean Honors List   09/20 – 12/23

MANUSCRIPTS & CONFERENCES

"A Reliable Change Estimation Method for Severely Cognitively Impaired Populations"
- **Manuscript** under revision for submission to *Alzheimer's & Dementia Journal*
- **Poster presentation** at Southern Regional Council On Statistics Conference (SRCOS)   06/25
- **Poster presentation** at Statistics in Pharmaceuticals Conference (SIP)   08/25
- **Poster presentation** at ASA NJ Statistics Workshop   10/25

"miRNA Expression in Breast Cancer Tissue and Plasma Pre- and Post-Neoadjuvant Chemotherapy: Associations with Hormone Receptor Status, Pathological Response, and Survival"
- **Manuscript** under submission to *World Journal of Clinical Oncology*

RESEARCH EXPERIENCE

Biostatistics Research Assistant   Providence, RI
supervised by Dr. Ani Eloyan, Assoc. Prof. & Vice Chair of Biostatistics, Brown University   09/24 – Present

——Project 1: Nonlinear Mixed Effect model for Alzheimer's Disease Biomarker   09/24 – 12/24
    (*Research Assistant contribution*)
- Conducted literature reviews on vivo Alzheimer's disease imaging biomarker (tau-PET, WMH in MRI) and Spline modeling techniques (B-spline, Natural-spline, penalized-spline)
- Constructed a P-spline design matrix based on cubic B-spline basis function within mixed-effects framework to model the nonlinear increment of White Matter Hyperintensities (WMH) volume in longitudinal MRI data.

——Project 2: A Bayesian Cognitive Change Method for Censored Longitudinal Data   01/25 – 07/25
    (*Independently led project; first author*)
- Designed simulation study and an R Shiny App to show that SRB method underestimates cognitive change in Severe dementia whose follow-ups are left-censored at floor of NACC test battery in Longitudinal EOAD Study (LEADs)
- Extended the SRB method by introducing Bayesian Inference that models censored follow-ups as random variables under a censored (Tobit-type) likelihood, realized through the MCMC (Hastings-Metropolis) implemented in R Stan.
- Conducted simulation and real-data analyses demonstrating the proposed method reduce the underestimation while preserving the interpretability aspect of the SRB index and providing credible intervals for uncertainty quantification.
- Resulted in a manuscript under submission to "Alzheimer's & Dementia Journal" and presentations at three conferences.

——Project 3: An Bayesian-based RNN Method for MNAR/MAR +left-censored longitudinal   09/25 – Present
    (*Independently led project; master thesis*)
- This thesis extends prior SRB-based work by replacing its predefined linear trend and single-retest constraint with an RNN architecture to fully utilize multiple follow-ups while preserving Bayesian uncertainty quantification.
- Constructed a unified likelihood that simultaneously handles MNAR/MAR missingness and left-censoring in Alzheimer's longitudinal data through RNN hidden states that capture temporal dependence.
- Implemented Bayesian inference, specifically variational inference, enabling the RNN to inherently model incomplete data while maintaining uncertainty quantification and computational efficiency.
- Basic derivations completed,with manuscript preparation in progress;; to be presented as a poster at Brown in April 2026

Translational Oncology Research collaboration   Providence, RI
collaborated with El-Deiry Laboratory, Legorreta Cancer Center, Brown University   03/25 – 09/25
- Led and conducted a full pipeline of statistical analyses for breast cancer miRNA biomarker discovery —including missing data imputation, survival data collection guidance, survival analysis.
- Identified MAR assumption via sensitivity analyses and applied conditional mean imputation for missing data
- Built Cox regression model to evaluate the predictive associations of miR-34a, miR-137, miR-373, miR-124a, and miR-155 with overall survival, taking account for both biological interpretability (up/down regulation, interactions) and statistical rigor (PH assumption, multicollinearity, influential point).
- Resulted in a manuscript under submission to the World *Journal of Clinical Oncology*.

Health Data Science Summer Fellowship   Providence, RI
supervised by Dr. Alyssa Bilinski, Assist. Prof. of Biostatistics, Brown University   06/25 – 08/25

- Received training in data tidying in R, visualization in Tableau, data management in SQL, and causal inference.
- Applied Large Language Model (LLMs) (OpenAI) for free-text classification of public response to private equity, aiming to evaluate and improve prompt effectiveness in domain-specific text understanding.
- Built an automated evaluation pipeline and interactive Shiny App to benchmark AI annotations against human ground truth.
- Resulted in oral presentation to program faculty and fellows at the final symposium.

Statistics Research Assistant
—Project 1: Graph-Based Change-Point Detection Method for High-Dimensional Network Data   09/22 – 01/23
Research Assistant contribution, supervised by Dr. Hao Chen, Assoc. Prof. of Statistics, UCD
- Conducted literature reviews on graph theory and graph models (Stochastic Block Models, Erdős–Rényi graphs and Configuration Model)
- Refined gSeg R-package, implementing a combination of Image Analysis and SCAN statistics test in detecting change-point in dynamic social networks, while optimizing the algorithms for computational efficiency and scalability
- Designed and conducted simulation studies in R to evaluate the predictive performance of the proposed method against conventional approaches, across different levels of graph sparsity, edge density, and network connectivity

—Project 2: Comparison of Machine Learning Models for Unemployment Prediction   12/22 – 03/23
Independent lead project, supervised by Dr. Colin Cameron, Distinguished Prof. of Economics, UCD
- Collected and managed unemployment related data through APIs, web scraping (BeautifulSoup) in Python
- Implemented and Compared multiple regression models, including OLS, LASSO, and a novel dimension-reduced local weighted regression applying PCA, for unemployment rate prediction.
- Evaluated and visualized model results in Stata and RStudio, using key metrics like Mean Squared Error (MSE), Mean Absolute Deviation (MAD), and R-squared, accuracy, and efficiency.

TEAM PROJECTS

Gun Violence Analysis Project   Davis, CA
Team Leader & Python Programmer   12/22 – 03/23
- Collected data using APIs and web scraping techniques (BeautifulSoup), and performed data manipulation in Python pandas
- Utilized Natural Language Processing (NLP) techniques, specifically NLTK, to explore the top 5 most frequently mentioned gun types concerning incidents of violence and causes in the violence with accurate lemmatization and stopwords removal
- Conducted visualization in python, including word cloud and interactive maps for gun violence exploratory data analysis

Airbnb Optimization Project   Davis, CA
Team Leader, Champion Project   06/22 – 09/22
- Directed a team in regression model development for Airbnb pricing, utilizing R for data processing and model selection.
- Validated models with K-fold Cross-Validation to prevent overfitting.
- Developed an R-based program that can optimize buyer location decisions by clustering house types
- Communicated results through LaTeX reports.

TEACHING & PROFESSIONAL EXPERIENCE

Master Teaching Assistant   08/25 – Present
- Hold office hours & Grade for "**Probability & Statistical Inference**" course at Biostatistics Department, Brown University

Biostatistician Summer Internship   Shijiazhuang, China
supervised by Dr. Jincai Hou, China Shineway Pharmaceutical Group Limited, National Laboratory   07/23 – 10/23
- Assisted statistical analysis for a Phase II clinical trial on vascular dementia treatment with a double-blind, placebo-controlled design, processing data across placebo (113 cases), low-dose (113 cases), and high-dose (114 cases) groups
- Applied survival analysis to evaluate treatment efficacy and time-to-event outcomes.
- Prepared reports and contributed to presentations for the drug development team

SKILLS

Programming: Python, R, Stata, SQL, MATLAB
Statistical Analysis: Generalized linear models, Longitudinal Data Analysis(LMM), Time Series Analysis(ARIMA), SurvivalAnalysis(PH, AFT), Categorical Data Analysis, Multivariate Analysis, Missing Data Analysis, Non-parametric & Parametric Test, Causal Inference, Bayesian Statistics, MCMC, Machine learning, Simulation Study, Sensitivity Analysis
Big Data Analysis: Web-scrap, PCA, Classification(logistics regression, LDA ,Regression Tree, Bagging, Random Forest, Boosting), Clustering (K-means, Hierarchical Clustering), Neural Network (RNN, CNN), Parallel Compute, LLM, High-dimensional Data

AWARDS

Health Data Science Summer fellowship / Brown University Scholarship / SRCOS Conference Travel Award Winner / UCDavis Deans' Honors Lists