Weijia Xiong

100 Haven Avenue, Tower 3, Apt.9H, New York, NY 10032 wx2233@cumc.columbia.edu

EDUCATION

Columbia University, Mailman School of Public Health, New York, NY expected May 2021
Master of Science, Biostatistics GPA: 3.92

East China Normal University, Shanghai, China June, 2019
Bachelor of Science, Statistics GPA: 3.6

University of California at Los Angeles, Los Angeles, CA Summer, 2017
Summer Program, Statistics

PROFESSIONAL EXPERIENCE

Roche Shanghai, China
Statistical Programming Analysis (SPA) Intern July, 2018 – Dec, 2018

- Participated in development of AutoReport RShiny app to streamline regulatory reporting process and integrate all related data files and documents in a one-stop shop; currently online and actively used.
- Completed design of GitHub web pages for team projects.
- Assisted team for with user acceptance testing (UAT), including questionnaire creation, data collection and final review.

RESEARCH EXPERIENCE

Columbia University Research New York, NY

Exploration of Super Learner algorithm Aug, 2020 – Present
Research assistant, advised by Dr. Caleb Miles

- Learned and reproduced the analysis of OTNY data, predicting three months of education enrollment or employment among unenrolled and unemployed at baseline via Super Learner algorithm.
- Applied Super Learner to Heart disease dataset and compared the predicted classification error rate with other machine learning methods (eg. GLM, Lasso, SVM, etc).

An application of Hidden Markov model for mobile mental health study June, 2020 – Present
Research assistant, advised by Dr. Linda Valeri

- Simulated outcome survey data and covariates (e.g. mobile-communication data like phones and calls, season, monthly clinical diagnostic scores) with mutual relations (e.g. AR (1), nonlinear pricewise function) between latent disease states.
- Created a mixed-effects Hidden Markov model with a multinomial logistic regression model for time-dependent covariates.
- Explored the causal effect of mobile-communication data and evaluated the prediction performance.
- Re-performed the model considering different missing data scenarios (MCAR, MAR, and MNAR) with different imputation methods (MICE, missForest, and median) or removing. Compared the results of prediction error rate.

Longitudinal microbiome data analysis Apr, 2020 – Present
Research assistant, advised by Dr. Jianhua Hu and Dr. Jian Wang

- Participated in a vaginal longitudinal microbiome research. Assisted in proposing a new cluster method for overall trajectory.
  - Imputed missing values via the KNN method.
  - Reproduced the clustering work for vaginal community state type and performed other methods (Hierarchical, K-means)
  - Conducted comparison of preterm birth classification for different methods (Lasso, Random Forest, SVM, etc).
- Participated in a longitudinal analysis of the microbiome in acute myeloid leukemia (AML) patients via Logistic Functional Additive Regression Method.
  - Performed data pre-processing and analysis for compositional data.
  - Implemented the algorithm to predict the infection of AML and assisted in building the package

Analysis of vehicle collision in New York City Sep, 2019 – Dec, 2019
Research assistant, advised by Dr. Jeff Goldsmith

- Analyzed NYPD motor vehicle collisions-crashes data with related covariates weather, traffic light, restaurant and holiday.
- Built the regression model and analyzed the correlation between potential factors and vehicle collisions.
- Created a website with vehicle collision map and shiny dashboard for data visualization and conclusions.

PUBLICATION

Bao, Y*, **Xiong, W***(2020). One-round communication efficient distributed M-estimation (submitted)

Bao, Y*, Liu, W*, Mao, X* and **Xiong, W***. Byzantine-Tolerant distributed multiclass sparse discriminant Analysis (manuscript in preparation)

TECHNICAL SKILLS

Computer skills: R, Matlab, LaTex, SAS, VS Code, Git