Yimeng Shang

☎ +86 15388260313 ✉ ys3298@cumc.columbia.edu https://ys3298.github.io/

EDUCATION

Pennsylvania State University, Hershey, PA — Aug, 2021 – May, 2025
Doctor of Philosophy in Biostatistics

Columbia University, Mailman School of Public Health, New York, NY — Apr, 2021
Master of Science in Biostatistics (GPA: 4.0/ 4.0)

East China Normal University, Shanghai, China — Jun, 2019
Bachelor of Science in Mathematics and Applied Mathematics

University of California, Berkeley, Berkeley, CA — Aug, 2017 – May, 2018
Berkeley International Study Program

INTERNSHIP EXPERIENCE

Eli Lilly China, Shanghai, China — Sep, 2018 – Jun, 2019
Data Science& Solution Intern

- Supported data management work in clinical trials including data cleaning and missing data query
- Constructed quantitative analysis of possible interference risks during clinical trials and organization operation,
- Developed an automatic web page with RShiny for reproducible monthly analysis to improve efficiency

RESEARCH EXPERIENCE

Variable Selection and Prediction in Logistic Regression with Misclassification in the Response
Research assistant, supervised by Professor Hua Shen, University of Calgary — May, 2020 – Dec, 2020

- Proposed a new method combining EM algorithm and Adaptive LASSO for variable selection, parameter estimation, and prediction. Extended the method to be feasible with fewer assumptions
- Constructed multiple simulation studies under different settings to evaluate the robustness of the new method
- Achieved consistent improvements in parameter estimation, variable selection and prediction under different settings
- A manuscript is prepared to submit for publication as the first author

Statistical Analysis of High Dimensional Metabolomics Data in ASD — May, 2020 – Aug, 2020
Research assistant, supervised by Professor Xiaoyu Che, Columbia University

- Constructed both logistic regression and Cox hazard model to estimate the effect size; adjusted for multiple comparison using Hochberg step up method; Did power analysis to compare the models
- Applied Bayesian generalized linear models to calculate credible intervals, select analytes with large Bayesian factors
- Implemented machine learning methods as feature selection methods to build a robust predictive model.
- A manuscript is prepared for publication with expected date of 2021

COVID-19 Psychological distress among healthcare workers at follow-up — Dec, 2020 – Mar, 2021
Research assistant, supervised by Professor Codruta Chiuzan, Columbia University

- Built longitudinal regression models to identify the risk factors for PTSD using longitudinal follow up data
- Used GEE to build longitudinal models with different correlation structure; Used Hosmer-Lemeshow test to check goodness of fit
- Applied factor analysis and penalized regression for dimension reduction and variable selection
- A manuscript is prepared for publication with expected date of 2021

TEACHING EXPERIENCE

Columbia University, Department of Biostatistics — Fall, 2020
Graduate Teaching Assistant, Course: BISTP8130 Biostatistical Methods I (95 students)

- Prepared homework and exam problem set and solutions; Graded homework and exams; Held weekly office hours;

SKILLS

Computer Skills: R, SAS, Python, MATLAB; **Language Skills:** Chinese Mandarin (Native), English (Fluent)