

Yimeng Shang

+1 (646) 704-5390

ys3298@cumc.columbia.edu

[personal website](#)

EDUCATION

Columbia University, Mailman School of Public Health, New York, NY
Master of Science in Biostatistics (GPA: 4.0/4.0, 4.08/4.3)

April, 2021

East China Normal University, Shanghai, China
Bachelor of Science in Mathematics and Applied Mathematics
Awarded ECNU Third-prize Academic Scholarship twice

June, 2019

University of California, Berkeley, Berkeley, CA
Berkeley International Study Program

August, 2017 – May, 2018

RESEARCH EXPERIENCE

Variable Selection and Prediction in Logistic Regression with Misclassification in the Response

June, 2020 – December, 2020

Research assistant, supervised by Professor Hua Shen, University of Calgary

- 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 by constructing two separate penalty term structure in EM algorithm and investigate misspecification problem
- Constructed multiple simulation studies under different settings, including considering different model assumptions, different dimensionalities, different regularization methods and different tuning parameter selection criteria to evaluate the robustness of the new method
- Achieved consistent improvements in parameter estimation, variable selection and prediction compared to the Naïve method and Ad Hoc method under different settings
- A manuscript is prepared to submit for publication as the first author

Statistical Analysis of High Dimensional Metabolomics Data in ASD

Summer, 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 and did sensitivity analysis by adjusting for potential confounding variables and testing interaction terms
- Applied Bayesian approach to calculate credible intervals and select analytes with large Bayesian factors
- Implemented Adaptive LASSO, Random Forest and XGBoosting algorithms as feature selection methods with bootstrap for the purpose of building a robust predictive model.
- A manuscript is prepared for publication with expected date of 2021

Functional Data Analysis for Mice Brain Electrophysiology Data

November, 2020 – Now

Research assistant, supervised by Professor Todd Ogden, Columbia University

- Built functional on scalar regression models and mixed effect models to picture the underlying pattern and association of brain electrophysiology curve under each condition
- Did functional permutation test to test the significance of the functional coefficients

COVID-19 Psychological Distress Among Healthcare Workers at Follow-up

December, 2020 – Now

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

ACADEMIC PROJECTS

COVID-19 Analysis via Logistic Curve and Clustering

April, 2020

Course: The Advanced Topics in Statistical Computing, Columbia University

- Estimated parameters of a logistic curve by first obtaining reasonable initial values with ordinary differential equations for gradient descent model to precisely estimate the parameters.
- Implemented both Gaussian Mixture Model with EM algorithm and K-means algorithm to cluster the estimated parameters for each region
- Used the clustering results drawing world map to visualize (e.g. drawing the word map with ggplot2) the distribution of clusters to find the potential geometric characteristics of COVID-19.

Dynamics of Ebola Transmission

October, 2018 – April, 2019

Independent Study, Advised by Professor Ping BI, East China Normal University

- Use R to produce time-based graphics indicating number of infections, death toll and increasing rate of infectious victims based on infection cases and death data in Guinea and Sierra Leone reported by WHO, and realize smoothing with splines
- Describe Ebola outbreak with SEIR dynamic system model, studied existence and stability of balance point with MATLAB and verify solution stability in the method of numerical simulation
- Propose suggestions for preventing Ebola virus spread based on analysis of parameter sensitivity

Statistical Methods in Medical Research and Development

August, 2018 – September, 2019

Independent Study, Advised by Professor Jin Xu

- Referred to medical research papers concerning medicines from top 10 pharmaceutical companies around the world and wrote summaries of the adopted statistical methods
- Attended weekly seminars related to statistical methods of medical journals for clinical trials at local hospital
- Reproduced a statistical analysis (ANOVA, chi-square test, t test) with given clinical data and compared the results with findings from the original statistical report

INTERNSHIP EXPERIENCE

Eli Lilly China, Shanghai, China

September, 2018 – June, 2019

Data Science & Solution Intern

- Supported data management work in clinical trials including data cleaning and missing data query
- Participated in the discussion of the application of statistical methods in clinical trials.
- 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

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; Attended lectures and answered questions

SKILLS

Computer Skills: R, SAS, Python, MATLAB

Language Skills: Chinese Mandarin (Native), English (Fluent)