Yimeng Shang

☎ +1 (646) 704-5390 ✉ yqs5519@psu.edu https://ys3298.github.io/

EDUCATION

Pennsylvania State Univerisity, Hershey, PA — August, 2021 - Present
Ph.D. in Biostatistics (GPA: 4.0/ 4.0)

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

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

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

INTERNSHIP EXPERIENCE

Cytel, Boston, MA, — June, 2022 – August, 2022
Strategic Consulting Intern

- Proposed predictive variable selection algorithm via knockoff filters
- Developed Shiny App to implement the proposed algorithm
- Supported clinical trial design simulation project and wrote the statistical analysis plan for FDA submission

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 under supervision of China DSS team
- 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

RESEARCH EXPERIENCE

Survival Analysis of After Transplant Leukemia Patients Relapse — Summer, 2022
Research assistant, supervised by Dr. Shouhao Zhou, Penn State University

- Applied time-varying competing risk model to find association between chimerism increase and non-relapse death
- Developed graphical representation and statistical test for irreversible time-dependent ordinal covariate for survival and competing risk analysis

Variable Selection in Multiple Regression with Misclassification in the Response Variable — Summer, 2020
Research assistant, supervised by Dr. 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 less 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 new method
- Achieved consistent improvements in parameter estimation, variable selection and prediction compared to the Naïve method and Ad Hoc method under different settings

Statistical Analysis of High Dimensional Metabolomics Data in ASD Summer, 2020

Research assistant, supervised by Dr. 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 generalized linear models 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.

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
- 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.

PUBLICATIONS

Endres K M, Kierys K, **Shang Y**, et al. A Multicenter Retrospective Evaluation of Specialized Laboratory Investigations in the Workup of Pediatric Patients With New-Onset Supraventricular Tachycardia[J]. Journal of Emergency Nursing, 2022.

Shechter, Ari & Chiuzan, Cody & **Shang, Yimeng** & Ko, Gavin & Diaz, Franchesca & Venner, Hadiah & Shaw, Kaitlin & Cannone, Diane & McMurry, Cara & Sullivan, Alexandra & Rivera, Reynaldo & Vose, Courtney & Shapiro, Peter & Abdalla, Marwah. (2021). Prevalence, Incidence, and Factors Associated with Posttraumatic Stress at Three-Month Follow-Up among New York City Healthcare Workers after the First Wave of the COVID-19 Pandemic. International Journal of Environmental Research and Public Health. 19. 262. 10.3390/ijerph19010262.

Abdalla, Marwah & Chiuzan, Cody & **Shang, Yimeng** & Ko, Gavin & Diaz, Franchesca & Shaw, Kaitlin & McMurry, Cara & Cannone, Diane & Sullivan, Alexandra & Lee, Sung A & Venner, Hadiah & Shechter, Ari. (2021). Factors Associated with Insomnia Symptoms in a Longitudinal Study among New York City Healthcare Workers during the COVID-19 Pandemic. International journal of environmental research and public health. 18. 10.3390/ijerph18178970.

SKILLS

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