Sibei Liu

sl4660@columbia.edu (646)361-2079 https://liu0223.github.io/Liu_Si_Bei.github.io/resume1.html

Education

Columbia University, Mailman School of Public Health, New York, NY *May,2021(Expected)*

Master of Science in Biostatistics (Theory and Method Track) GPA:3.75

Jilin University, School of Public Health, China *June,2019*

Major: Information Management & Information System

Honors: 05/2018 China Scholarship Council- Mitacs (Canada) Summer Internship funding

09/2017 China National Scholarship in Academic Year 2016-17 awarded by Ministry of Education of China (0.2%)

Related Courses and Skills:

Courses: Probability, Biostatistical Methods (regression modeling, survival analysis, longitudinal analysis), Data Science 1, Data Science for Machine Learning (Regularization, non-linear regression, Classification), Statistical Inference, Epidemiology, Advanced Statistical Computing (Monte Carlo, Optimization Algorithms, Bootstrap, MCMC), SAS (Observer)

Skills: Good command of R, SPSS, Microsoft Office (Access, Excel, Word, PowerPoints), SAS (Base Certificate)

Languages: English, Mandarin (Native)

Research Experience

Machine Learning:

Classification on autism status using structural MRI measures *Research Assistant* *July, 2020- Current*

New York University, New York, NY

- Cleaned dataset and built a logistic-Lasso model to select predictors
- Used selected variables to build tree model and applied ensemble methods to check sensitivity and specificity

Meta-Analysis and Systematic Reviews:

Meta-Analysis Projects *Research Assistant* *November, 2019 - January, 2020*

Columbia University, New York, NY

- Used random effects model to find the pooled mean differences or ORs and produced forest plots in R
- Examined sources of between-study heterogeneity
- Gave presentations to dental students on Doing Meta Analysis in R

Child Maltreatment and Genetic Variation in Adult Psychopathology *Research Assistant* *July, 2018- October,2018*

McGill University, Montreal, Canada

- Extracted available data from publications and completed quality evaluation for screened papers
- Divided and summarized quantitative and qualitative papers
- Wrote the manuscript, which has been published

Risk Factors Exploration:

Sociodemographic and Clinical Predictors of Myelitis Encephalitis *Research Assistant* *May,2020 – July 2020*

Columbia University, New York, NY

- Outputted the descriptive tables of social and clinical covariates
- Imputed the missing data using and did logistic regressions to explore risk predictors
- Wrote the Statistical Method part manuscript

The Influencing Factor of Teenager Smoking in Liaoning Province *Research Assistant* *March, 2018-May,2018*

Jilin University, Changchun, China

- Determined the innovative perspective of data analysis
- Used the complex sampling pack in SPSS to analyze multistage stratified cluster sampling, including completing Chi-square analysis and the logistic analysis to find risk factors of teenager smoking

Team-worked Class Projects

Columbia University, New York, NY

Bayesian Model for hurricane speed *May, 2020*

- Designed an MCMC algorithm to estimate the posterior distribution of model parameters (no package used)
- Estimated parameter using posterior mean and 95% credibility interval and evaluated the prediction performance

Daily COVID-19 Cases Prediction model Across Nations *April, 2020*

- Estimated the coefficients of logistic curve model by using the Adam algorithm (no package used) and visualized data
- Used Mixed Gaussian Model performed by EM algorithm to cluster the estimated parameters of each nation

Optimization on breast cancer data *March, 2020*

- Use Newton-Raphson to estimate of the coefficients logistic regression model (no package used)
- Built a logistic-Lasso model to select features and used Path-wise Coordinate-wise optimization to find a path of solutions with a sequence of lambdas (no package used)
- Used Cross-validation to select the best lambda with highest AUC

Simulation study to compare three survival models *February, 2020*

- Used Monte Carlo algorithm to generate 3 sets of survival data with fixed true parameters which follow Exponential proportional hazard model, Weibull proportional hazard model and Cox proportional hazard model respectively (no package used)
- Fitted three models under each set of data and simulated it for 10000 time a set of betas were extracted and used to calculate the mean and 95% CI to do the comparison with true parameter

The influencing factors of U.S. Suicide Rate *November, 2019*

- Extracted data from 11 website, produced world suicide graphics by using R
- Performed multiple linear regression to find the influencing factors of suicide
- Built a github-based website to show the project report（https://ys3298.github.io/life_saver.github.io/index.html）

Publications

- Pallavi Misra, ***Sibei Liu***, Xiangfei Meng, What DNA methylation modifications and/or genetic variations interact with childhood maltreatment in the development of depression: A systematic review, Journal of Affective 252,2019, 325-333
- Tingting Gao et al, including ***Sibei Liu***, The Role of School Connectedness and Maladaptive Cognitions in the Association between Stress and Internet Addiction: A Serial Mediation Model, Perspective Psychiatric Care. 2019; 55: 728‐733
- Muzi Li, ***Sibei Liu*** et al, Interactions of childhood maltreatment and genetic variations in adult depression: a systematic review[J]. Journal of affective disorders, 2020.