Zhengwei Song

Tel: (332) 256-6895 | Email: zs2539@cumc.columbia.edu | Website

EDUCATION

Columbia University Mailman School of Public Health, NY, USA Sep 2022 – May 2024 (Expected)

MS Biostatistics (GPA: 3.96)

- Relevant Coursework: Probability, Biostatistical Methods, Epidemiology, Data Science, Survival Analysis, Randomized Clinical Trial, Data Mining, Advanced Statistical Computing, Advanced Probability

University of Manchester, Manchester, UK Sep 2019 – Jul 2021

BSc Mathematics and Statistics (GPA: 3.65, 1st in class)

- Relevant Coursework: Real Analysis, Statistical Inference, Markov Chain, Martingales

Shandong University, Jinan, China Sep 2017 – Jun 2021

BS Mathematics (GPA: 3.62, top 30%)

- Relevant Coursework: Advanced (Linear) Algebra, Calculus, Geometry

WORK EXPERIENCES

Teaching Assistant, Columbia University (New York) Sep 2023 – Dec 2023 (Expected)

- Hold weekly office hours, grade homework, attend the instructor's flipped classroom for Q&A
- Two courses: Mathematical Statistics, Biostatistical Method I

Biostatistics Intern, Medical Scientific Affairs Dept, Roche Diagnostics (Shanghai) Apr 2022 – Sep 2022

- Developed statistical methods for analyzing clinical trial data, including the development of novel approaches to address specific research questions and issues with existing methodologies
- Collaborated with the medical team and provided statistical support in Phase IV clinical trials by developing analytical plans, performing analyses, interpreting results, and summarizing findings into concise reports that are understandable to non-statisticians
- Developed an R package (jmpost) of linear mixed effects models for the tumor size over time by Bayesian inference using MCMC

Data Analyst Intern, Information System Dept, Sina Weibo (Beijing) Oct 2021 – Apr 2022

- Responsible for scraping & wrangling user data, and creating visualization (user portraits) for rankings in the entertainment operations, and presented final statistics for several popular TV series, variety shows, and documentary, to provide data support for social media influencers and internal operations
- Maintained Hive SQL and data warehouse services

RESEARCH EXPERIENCES

Multiomics in Alzheimer's Disease (AD) Mar 2023 – Present

Annie Lee Lab, Dept of Neurology, Columbia University

- Perform the adaptive gene-environment interaction test to find significant genes
- Implement mediation analysis across pathologies, gene expression, infarcts and AD
- Conduct new cluster analyses (SNF, CCA, etc.) across different cohorts, identifying subtypes of ageing individuals with differences in Alzheimer's risk and the impact of top significant genes

Data Simulation by Multilevel Monte-Carlo for Black-Scholes Pricing Model Sep 2020 – Jun 2021

Kody Law Lab, Dept of Mathematics, University of Manchester

- Applied the Euler-Maruyama Method to solve the discretized linear stochastic differential equations on Ito Integral version

- Implemented Monte-Carlo simulation of expectations with strong and weak rates of convergence
- Deduced the properties of an independent estimator by a sequence of Monte-Carlo samples
- Applied Rejection Sampling to Black-Scholes pricing model and improved computational efficacy

Edible Tableware based on Finite Element Analysis Apr 2019 – Apr 2020
Song Yu Lab, School of Mathematics, Shandong University

- Led a team with diverse academic backgrounds and secured full funding (around $900)
- Designed and produced a chopstick-like mold, testing its mechanical characteristics
- Connected and partnered with local restaurants and pubs for testing and review

PUBLICATIONS

"Genome-wide gene-based study in multi-ethnic cohorts identifies genes that interact with vascular risk factors in Alzheimer's disease (AD)" (with Annie Lee), under review for *Alzheimer's disease and Parkinson's disease Conference,* 2024.

SKILLS

Computer

- R: tidyverse, ggplot2, caret, survival, bioconductor, lme4, gee, httr, etc.
- SAS, RShiny, C, MATLAB, SQL, Unix, Microsoft Office, AutoCAD

Tests

- Continuous: t, z, ANOVA
- Categorical: chi-squared, Fisher's exact, McNemar's
- Non-parametric: sign, Wilcoxon signed-rank and rank-sum, Log-rank

Modeling / Machine Learning

- Linear: linear, glm (e.g. logistic, Poisson), weighted least squares
- Longitudinal: mixed effect, GEE, survival (Cox, Stratified PH, AFT)
- Model Selection: LASSO, Elastic net, Ridge, PCR, PLS, cross validations
- Non-parametric: decision tree, random forest, boosting, K-NN, cubic splines, local regression, GAM, MARS, LDA, QDA, NB, SVM, etc.
- Unsupervised: clustering (K-means, Hierarchical, spectral), PCA, etc.

Simulation & Optimization

- Data generation, Newton-Raphson, EM, bootstrapping, MCMC

AWARDS

AAAI 2022 Security AI Challenger VIII Award (44 out of 3692) Jan 2022

Enactus Annual Outstanding Individual (Top 10%) Dec 2018

Shandong University Third-level Academic Scholarship, 2017-2018 (Top 30%) Oct 2018

EXTRACURRICULAR ACTIVITIES

Treasurer, Enactus, Shandong University Oct 2018 – Oct 2019

- Tutored students in budget planning for every program
- Responsible for the entire budget arrangement of the organization
- Award: Enactus Annual Outstanding Individual (Top 10%)

Secretary, Association of International Exchange, Shandong University Oct 2017 – Oct 2018

- Assisted in events and activities for the international student exchange
- Trained students to avoid and bridge the gap of culture shock