

JUNZHE SHAO

Columbia University, Mailman School of Public Health

(+1)858-766-1048 \diamond js5959@columbia.edu \diamond <https://junzheshao5959.github.io/>

EDUCATION

Columbia University, Mailman School of Public Health Sep 2021 - Present
M.S. in Biostatistics

Peking University, School of the Life Sciences Sep 2016 - June 2021
B.S. in Biological Science
Minor in Physics

University of California, San Diego Sep 2019 - Sep 2020
Visiting Student

RESEARCH EXPERIENCE

Causal Inference for Non-Stationary Time Series Data (On going)

November 2021 - Present

Advisor: Dr. Linda Valeri, Dr. Xiaoxuan Cai, Dr. Mingzhang Yin, Dr. David Blei

- Proposed a novel parametric generalization of the synthetic control method with time-varying weight that is not permutation invariant. (working paper in progress)
- Demonstrated the model's flexibility using a state-space model, a Kalman filter/smoothing estimator for a time-varying weight, and using Bayesian hierarchical shrinkage for two-way sparsity across donor pool and time.
- Using this new approach to investigate the impact of a mandatory certificate on COVID-19 vaccine compliance.
- Developing an R package `SSMimpute` implementing the missing data imputation approach for "State space model multiple imputation for missing data in non-stationary multivariate time series".

Covariates Adaptive Design for Randomized Clinical Trial under High-dimensional Setting (On going)

June 2022 - Present

Advisor: Dr. Jingshen Wang

- Investigated the theoretical properties of statistical methods to achieve covariates balance for randomized clinical trials.
- Proposed a new setting for Pairwise Sequential Randomization with many covariates.
- Proved that a balancing criteria can be achieved asymptotically when the number of covariates is a vanishing proportion of the number of patients.
- Simulation shows better efficiency of pairwise sequential randomization compared to complete randomization with many covariates.

Integrative High-throughput Metabolomics Analysis of Pulmonary Arterial Hypertension Phenotypes and Outcomes

Sep 2019 - June 2021

Advisor: Dr. Mohit Jain, Dr. Tao Long

Publication: Alotaibi, Mona, Junzhe Shao, Michael W. Pauciulo, William C. Nichols, Anna R. Hemnes, Atul Malhotra, Nick H. Kim et al. "Bioactive Metabolomic Profiles of Scleroderma-PAH are different than idiopathic PAH and associated with worse clinical outcomes." CHEST (2022)

- Built a regularized regression statistical model based on high throughput mass spectrum of bioactive lipids in plasma sample to do mortality prediction of Pulmonary Arterial Hypertension (PAH).
- The model outperformed traditional clinical variants in the metric of AUC(0.78). Gave out a quick and non-invasive mortality risk score to be practically utilized.
- Further studied the properties of metabolites by molecular networking and the causal inference by Mendelian randomization.
- Adapted the model on the subtype prediction of PAH type I, using scleroderma and IPAH as the new response.

Image Based Age and Life Expectancy Prediction of C.elegans

Mar 2021 - Present

Advisor: Dr. Jingdong Jackie Han

- Developed a deep learning method based on Inception-ResNet-V2 of image processing to identify the movement of *Caenorhabditis elegans* model from microscopic video of different levels of resolution.
- Proposed a novel Multi-task learning approach for both prediction of mortality and life expectancy to get generalization ability.
- Defined features of vitality index and calculated frailty score for predicting the chronological age and life expectancy of worms.
- Achieve a Mean Absolute Error of 1.8 days for age results and 2.6 days for lifespan results.

RELATED COURSES

Mathematics Courses: Mathematical Analysis, Linear Algebra, Mathematical Methods for Physics (Complex Analysis and PDE)

Statistics Courses: Probability and Mathematical Statistics, Applied Regression Analysis, Casual Inference (PhD-Level), Applied Causality (PhD-Level) , Data Science I (Data Analysis Using R), Data Science II (Statistical Learning), Biostatistical Method I (Statistical Inference and Linear Regression), Biostatistical Method II (Generalized Linear Regression and Longitudinal Data Analysis), Applied Stochastic Process, Data Analysis of Genomics, Mathematical Modeling in the Life Sciences

Physics Courses: Theoretical Mechanics, Electrodynamics, Equilibrium Statistical Physics, Quantum Mechanics, Solid State Physics

EXTRACURRICULAR ACTIVITIES

Piano Club, PKU

June 2016 - Present

Beijing, China

Played Lindraja by Debussy in Duo Piano Concert of Peking University, 2017

SKILLS AND TECHNICAL STRENGTHS

Programming Languages

R, Python, MATLAB, C

Software & Tools

L^AT_EX, git, shell, Bash

Languages

Mandarian(Native), English(Proficient)