Wenxing (Blue) Wang

(203) 278-3035 | blue.wang@yale.edu

EDUCATION

Yale University   **New Haven, CT**
Master of Science in Biostatistics   *Sep 2021 – May 2023*
- **Relevant Coursework:** Data Structure (A-), Linear Models (H), Longitudinal and Multilevel Data Analysis (H), Machine Learning (H), Nonparametric Statistics (H), Optimization and Computation (H), Stochastic Processes (H), Reinforcement Learning (H), Theory of Survival Analysis (H)

Hong Kong Polytechnic University   **Hong Kong**
Bachelor of Science (Hons.) in Applied Biology with Biotechnology   *Sep 2017 – May 2021*
- **Honors & Awards:** Dean's Honor List 2018/19 (Top 5% of class), Dr. Y.K. Ching Memorial Scholarship, Subject Prize for Outstanding Performance 2018/19, 2019/20

Minor in Applied Mathematics   *Jan 2019 – May 2021*
- **Relevant Coursework:** C++ Programming (A+), Calculus and Linear Algebra (A+), Linear Regression (A), Probability Theory (A+)**,** Real Analysis (A+), Statistical Inference (A)

North Carolina State University   **Raleigh, NC**
Academic Exchange   *Jan 2020 - Jul 2020*

RESEARCH EXPERIENCE

North Carolina State University   **Raleigh, NC**
Advisor: Prof. Arnab Maity   *Mar 2020 – Present*
- Designed a novel hypothesis testing method on parameters of a regression model with Tukey's 1-degree-of-freedom form of interaction for testing possible gene-gene and gene-environment interactions when estimating genetic effects of phenotypes
- Derived the test statistic and proposed a permutation test that permutes residuals under null hypothesis; Implemented the mathematical derivation in R
- Verified the asymptotic properties of the test statistics and robustness of the proposed hypothesis testing by conducting simulation studies

Yale University   **New Haven, CT**
Advisor: Dr. Emma Zang   *May 2023 – Present*
- Explored the relationship between care recipient health and caregiver health over time by conducting group based Bayesian trajectory analysis
- Conducted comprehensive analysis of overall health ratings over time for paired care recipient and caregiver to identify trajectory trends and investigated population characteristics and transition probabilities to understand dynamic relationships in health outcomes
- Explored the causal effect of 1990 clean air acts (CAA) policy amendment on county-level mortality rate of older adults over 65
- Applied two-way fixed effect event study approach to examine the effectiveness in reducing mortality rates through the amendment of CAA and identified significant reductions in county-level mortality rate after the 1990 CAA amendment

Yale University   **New Haven, CT**
Advisor: Prof. Hongyu Zhao   *May 2023 – Oct 2023*
Link: https://github.com/bluestme/SDPRX_CPP
- Developed SDPRX, a hierarchical non-parametric Bayesian model software, in C++ for cross-population prediction of complex traits in genetics between 2 populations
- Conducted full-cycle software implementation, including data preprocessing and MCMC sampler, optimizing computational efficiency with multi-threaded coding and custom data structures
- Enhanced polygenic risk score (PRS) predictions, demonstrating a tenfold increase in computation speed compared to previous Python implementations, and published the tool on GitHub

Yale University   **New Haven, CT**
Advisor: Prof. Hongyu Zhao   *Jul 2023 – Present*
- Led and developed a pipeline to explored elusive cell types from multi-modality CITE-seq data
- Designed and implemented a novel diversity assessment metric based on the Local Inverse Simpson Index (LISI) for more accurate analysis of cell clusters in unbalanced datasets, validated through extensive simulation studies
- Tested the performance of modality integration tools using the discovered elusive cell types

Yale University   **New Haven, CT**
Advisor: Prof. Hongyu Zhao   *Jan 2024 – Present*
- Developed SDPR3, a hierarchical non-parametric Bayesian model software, in C++ for cross-population prediction of complex traits in genetics between 3 populations
- Conducted full-cycle software implementation and simulation studies with Plink2 and gcta64 to verify the performance of the model with various sample sizes, causal SNP percentage and cross-population correlations

SKILLS
- **Programming Languages:** R, Python, C, C++, SAS, Stata, SQL
- **Other:** Latex, Linux, Github, Microsoft Office