Mengshan Wu

Tel: 203-909-2333 | **Email**: mengshan.wu@yale.edu | Home Page

EDUCATION

Yale University — **New Haven, CT, USA**
Master of Science in Biostatistics — *Aug 2024 - May 2026*
- *GPA: 4.0/4.0*
- Relevant courses: Probability Theory (4.0/4.0), Theory of Statistics (4.0/4.0), Advanced Regression Models (4.0/4.0), Longitudinal & Multilevel Analysis (4.0/4.0), Bayesian Statistics (4.0/4.0), Machine Learning and Causal Inference (4.0/4.0), Applied Survival Analysis(4.0/4.0), Statistical Inference (ongoing)

Central University of Finance and Economics — **Beijing, China**
Bachelor of Economics in Finance — *Sept. 2020 - Jun 2024*
- *GPA: 91.69/100, rank top 5%*
- Relevant courses: Advanced Mathematics (95/100), Probability and Statistics (96/100), Statistics (99/100), Modern Statistics Software (95/100), Advanced Algebra (86/100), Operations Research (100/100)

University of California at Los Angeles — **Los Angeles, CA, USA**
Summer Session — *Aug 2023 – Sept. 2023*
- *GPA: 4.0/4.0*
- Relevant Courses: Real Analysis (4.0/4.0), Statistics and Data Mining (4.0/4.0)

RESEARCH INTEREST

- **Biostatistics**: Statistical methods for genomics and genetics; Multivariate Survival Analysis
- **Statistics**: High-dimensional Statistical Inference; Tensor Time Series

PUBLICATIONS AND MANUSCRIPTS

[1] Mou, Y.*, **Wu, M.***, Huang, Y., Hummel, S. Rotation Win Ratio: A Flexible Prioritization Method for Hybrid-Priority Composite Endpoints. Manuscript in preparation.

[2] **Wu, M.**, Zhao, J., Zhao, H. Evaluating methods for detecting overall and cell-type-specific spatially variable genes in spatial transcriptomics data. Manuscript in preparation.

[3] Chen, H.-Y., Guo, R., Hung, C.-C., Lin, Z.-H., & **Wu, M.** Behavioral Intentions of Bank Employees to Implement Green Finance. *Sustainability*. 2023; 15(15):11717. [Link]

RESEARCH EXPERIENCE

Time-varying Tensor Factor Model in CP Form for High Dimensional Tensor Time Series, with an Application of Airline Passenger Transport Volume Series
*Supervisor: **Prof. Rong Chen**, Department of Statistics, Rutgers University* — *July 2025 - present*
- Constructed a monthly tensor time-series dataset of U.S. airline passenger flows using public aviation records.
- Analyzed the dataset using Tucker and CP tensor factor models. The Tucker model required too many factors and produced diffuse latent patterns, while the CP model revealed interpretable hub structures, regional clusters, and carrier-specific dynamics, yet also indicated temporal structural changes not captured by standard CP frameworks.
- Developing a time-varying CP tensor factor framework with smoothly evolving factor loadings to accommodate gradual structural change and potential structural breaks in high-dimensional systems.

Benchmarking Methods to Detect overall and cell-type-specific Spatially Variable Genes
*Supervisor: **Prof. Hongyu Zhao**, School of Public Health, Yale University* — *Apr 2025 – present*
- Designed a controlled single-cell spatial transcriptomics simulation framework that integrates scCube to generate reference-based spatial coordinates with realistic clustered patterns, and supports a multi-factor experimental design with configurable cell-type compositions and signal-strength levels for evaluating ct-SVG detection methods.
- Benchmarked 5 SVG and ct-SVG detection methods (CELINA, SPARK, SPARK-X, MCube, STANCE) for single-cell and spot-level spatial data by evaluating type-I error under strictly null simulations and assessing power under different effect-size settings and cell-type proportion regimes across multiple FDR thresholds.
- Interpreted performance based on each method's underlying assumptions and statistical modeling framework.

Rotation Win Ratio Methodology for Composite Endpoint Survival Analysis
*Supervisor: **Prof. Yuan Huang**, School of Public Health, Yale University* — *Apr 2025 – present*
- Identified the methodological gap in Win Ratio works of endpoints sharing equal clinical importance, and proposed the Rotation Win Ratio (RWR) as a hybrid prioritization approach enabling multi-endpoint survival comparisons.

- Implemented U-statistics–based inference for RWR, including variance estimation, confidence interval construction, and global FS-type tests, and integrated the procedures into a unified computational framework.
- Developed a full R software package supporting rotation generation, pairwise comparison with censoring, stratified analyses, and joint covariance estimation under both null and alternative hypotheses.
- Designed and executed large-scale copula-based multivariate survival simulations (10000 replications across 15+ scenarios) to assess type-I error, power, and CI coverage across varying dependence structures, treatment effects, and censoring, demonstrating the robustness and interpretability of RWR.

Bayesian Graphical Modeling of Gene Co-expression Networks from scRNA-seq

*Supervisor: **Prof. Joshua Warren**, School of Public Health, Yale University* *Sept. 2024 – Dec 2024*

- Developed a Bayesian Poisson log-normal framework for cell-type-specific gene co-expression networks inference, incorporating sparsity via Wishart-MCMC and graphical LASSO priors within a block-wise Gibbs sampler.
- Applied Polya-Gamma augmentation to enable efficient posterior updates under a Poisson measurement model and conducted simulations evaluating sparsity recovery and robustness to sequencing noise.
- Applied the framework to 3,641 microglia nuclei from a PNAS dataset, generating biologically interpretable conditional-dependence networks supported by a reproducible R pipeline.

Predictive Modeling and Risk-Factor Analysis for Heart Disease

*Supervisor: **Prof. Yuehan Yang**, School of Statistics and Mathematics, CUFE* *Sept. 2022 – Jan 2023*

- Built a principled statistical modeling pipeline for heart-disease prediction and risk-factor analysis, with BIC-optimized Probit model achieving strong predictive accuracy (AUC~85%) while maintaining interpretability.
- Validated model robustness through baseline comparisons with Logit/Probit GLMs using AIC/BIC selection, and through consistency checks against regularized and tree-based machine learning models.

Work Experience

Baidu Online Network Technology Co., Ltd. **Beijing, China**

AIGC Data Analysis Intern, Strategy and Business Intelligence Group *Jun 2024 - Aug 2024*

- Built automated analytics and predictive models for retention and engagement using large-scale AIGC user data.

China International Capital Corporation Limited **Shanghai, China**

Quantitative Strategy Research Intern, Financial Engineering Group *Sept. 2023 - Nov 2023*

- Analyzed sub-IPO time series to build and validate risk-factor models capturing time-varying return dynamics.

National Institute of Financial Research, Tsinghua University **Beijing, China**

Research Intern, Green Finance Research Center *Mar 2023 - Jun 2023*

- Built a CGE-based model to assess macroeconomic and default-risk impacts of low-carbon transitions.

Beijing Jian FinTech Co., Ltd **Beijing, China**

Intern, Fund Research Department *Jul 2022 – Sept. 2022*

- Analyzed corporate financial and ESG data to develop governance metrics and benchmark firm performance.

Teaching Experience

- Yale University - S&DS 2420: Theory of Statistics, TA *2026 Spring*
- Yale University - BIS 567: Bayesian Statistics, TA *2025 Fall*
- Central University of Finance and Economics - Probability and Statistics, TA *2022 Fall*
- Central University of Finance and Economics - Students' Mental Health, TA *2021 Fall*

Awards and Honors

Scholarship

- Chengwang Scholarship, CUFE ($7,500) *Jun 2024*
- Outstanding Scholarship for Academic Excellence, CUFE (2%) *Oct 2023*
- Outstanding Scholarship for Academic Research and Innovation, CUFE (2%) *Sept. 2022*
- First-class Scholarship, equals to Dean’s List, CUFE (2%) *Sept. 2022*

Contest

- Champion of Singing Competition, CUFE / Yale *2025, 2023*
- “Internet+” Innovation and Entrepreneurship Competition, First Price of Beijing *Oct 2023*
- College Student Innovation and Entrepreneurship Training Program, National Excellent Project *Jul 2022*

Skills

Programming R, Python, PyTorch, Matlab, SAS, MySQL, Stata

Platforms Linux, GitHub, HPC/Slurm