Yingzhen Wang

yingzhen19970316@gmail.com | (734) 369-0963 | www.linkedin.com/in/yingzhen-wang

EDUCATION

University of Michigan   **Ann Arbor, MI**
Master of Science (MS) in Biostatistics | GPA: 3.860/4.0   August 2024 - Present
Completed Coursework: Probability & Distribution Theory, Theory & Application of Linear Regression, Computing with Big Data, Introduction to Epidemiology, Analysis of Categorical Data, Theory and Application of Generalized Linear Models, Biostatistics Inference.
Current Coursework: Programming Laboratory in Bioinformatics, Clinical Trials, Theory and Application of Longitudinal Analysis.

Colorado State University   **Fort Collins, CO**
Master of Applied Statistics, Statistical Science Specialization | GPA: 3.904/4.0   August 2020 - August 2022
Relevant Coursework: Machine Learning, Statistical Learning & Data Mining, Data Visualization Methods, Applied Bayesian Statistics, Quantitative Reasoning, Generalized Regression Models, R Software, etc.

Yanbian University   **Yanji, China**
BS in Statistics | AVG: 81.39/100   September 2016 - June 2020
Honors: Departmental Scholarship (2018, Top 15%), Excellent Student Cadre (2017-2019, Top 5%)
Relevant Coursework: C Language, Multivariate Statistical Analysis, Regression Analysis, Stochastic Process, Mathematical Modeling, Statistic Software, Ordinary Differential Equation, Algebra & Geometry, etc.

PUBLICATIONS

[1] **Y. Wang**. 2024. "Research on Predictive Algorithms for Cardiovascular Disease." *The 2023 4th International Symposium on Artificial Intelligence for Medicine Science (ISAIMS '23)*. Association for Computing Machinery, New York, NY, USA, 304–314. https://doi.org/10.1145/3644116.3644169

[2] Z. Zhang, J. Niu, and **Y. Wang**, "The Application of Principal Component Regression Based on the Standard Model of Subsistence Allowance-Taking Jilin Province as an Example," *Economic Outlook the Bohaisea*, pp. 88, 2019, doi: 10.16457/j.cnki.hbhjjlw.2019.02.064.

[3] Z. Zhang, **Y. Wang**, and W. Sun, "Application of Nonlinear Programming to Heat Conduction Model," *International Journal of Engineering and Management Research*, vol. 8, no. 5, pp. 169-172, 2018, doi: doi.org/10.31033/ijemr.8.5.06.

RESEARCH EXPERIENCE

Research on Pneumonia Patient Condition Classification Using Diffusion Models and CLIP Ann Arbor, MI
Research Assistant   Nov 2024 - Dec 2024
- Deployed the stable Diffusion model and utilized LoRA to fine-tune the model to make the model generate customized chest X-ray images.
- Utilized the fine-tuned contrastive language-image pre-training (CLIP) model to achieve modest improvements in training accuracy and classify the test dataset.
- Publish the whole research on GitHub *https://github.com/xxm12345666/biostat625-group2-project.git*.

Research on Predictive Algorithms for Cardiovascular Disease   **Remote**
Summer Research Seminar, Supervisor: R. Todd Ogden, Columbia University   May 2023 - July 2023
- Utilized Principal Component Analysis (PCA) to identify relevant predictors and reduce dimensionality in complex datasets, implemented backward stepwise elimination to refine model features and prevent overfitting.
- Developed random forest and logistic regression models to assess CHD risk factors and aid in early detection.
- Evaluated model performances based on confusion matrix and AUC-ROC curve values.
- Validated that the random forest model outperformed other models in categorizing high-dimensional data.

- Accepted by *the 2023 4th International Symposium on Artificial Intelligence for Medical Sciences (ISAIMS 2023)* and published on *Association for Computing Machinery*.

Familial Influences on Radiation Effects in Mice   **Fort Collins, CO**
NASA Human Research Program, Weil Lab, Colorado State University   May 2022 – August 2022
- Designed and fitted a generalized linear mixed model for the analysis of the relationship among Modified Merriam-Focht classification and radiation groups.
- Detected the difference in radiation effects between γ-rays and HZE nuclei using emmeans command.
- Achieved regression analysis for the odds of being vision impairing caused by gamma rays using R.

Application of Nonlinear Programming to Heat Conduction Model   **Yanji, China**
Research Assistant   August 2017 - October 2018
- Established a nonlinear programming model to determine optimal thickness of the second layer of the high temperature working clothes given predictors like work hours and environmental temperature.
- Obtained the function of temperature and material thickness based on Fourier heat law.
- Introduced the simulated annealing algorithm to enhance optimization accuracy, considering convection and radiation factors.
- Authored a paper and published on *International Journal of Engineering and Management Research.*

The Application of Principal Component Regression Based on the Standard Model of Subsistence Allowance-Taking Jilin Province as an Example   **Yanji, China**
Research Assistant   September 2018 - February 2019
- Collected data of six main indicators of subsistence allowance in Jilin Province from 2010 to 2016 based on China Statistical Yearbook, calculated their correlation coefficients with MATLAB to validate the feasibility.
- Performed PCA with R to establish models of urban and rural subsistence standards.
- Obtained a more reasonable standard model of subsistence allowance by fitting the relationship between the six variables and time based on Gaussian function.
- Authored a paper and published on *Economic Outlook the Bohaisea.*

COMPETITIONS & INTERNSHIP
China Everbright Bank, Changchun Branch   **Changchun, China**
Manager Assistant   September 2019 - November 2019
- Gathered and processed the monthly economic data of Jilin Province and the whole country since 2019 to understand the GDP drivers from the perspectives of fiscal policy and monetary policy.
- Learned the business models and daily operations of commercial banks; recommended appropriate wealth management products to potential clients based on their risk tolerance and investment needs.
- Employed R to fit generalized and linear regression models for customer satisfaction and deposit analysis.

Research on Migration Patterns of Herring and Mackerel   **Yanji, China**
Team Leader, The 6th Yanbian University Mathematics Modeling Contest   April 2019
- Led a team of three in building a particle swarm model to make a reasonable judgment of the migration of herring and mackerel.
- Gathered the data of Atlantic Ocean temperatures from 1970 to 2019, built an Arima Model and Time Series Model in MATLAB and R to forecast the monthly average temperature of each latitude and longitude for the next 50 years.
- Applied particle swarm optimization (PSO) algorithm to build fitness function, which could locate the location where fish migrate over different months using data on water temperature.
- Proposed suggestions to ensure sustainability of fish population and economic profit and won the Second Prize.

Research on gene expression in the mouse brain using SPADE for Alzheimer's disease   **Ann Arbor, MI**
Research Assistant, Department of Computational Medicine & Bioinformatics   May 2025 - July 2025
- Conducted spatial transcriptomics analysis on normal vs. Alzheimer's mouse brains using SPADE, identifying over 13,000 spatially variable genes with high sensitivity to local expression heterogeneity.
- Performed clustering and domain detection with the BASS package and integrated Seurat-based differential expression analysis to compare spatial domain-specific vs. localized gene expression.

- Demonstrated that SPADE is good at detecting fine-scale localized differences, while Seurat capturing domain-level transcriptional changes, providing complementary insights into Alzheimer's-related gene dysregulation.
- Built reproducible workflows on Great Lakes HPC with R&Python environments, executing Slurm jobs for the 3000*30000 expression matrix through VS Code.

Research on Bayesian Parameterization   **Ann Arbor, MI**
Research Assistant, Department of Biostatistics   May 2025 - Present
- Designed and implemented elliptical slice sampling algorithms in R for Bayesian parameter estimation, improving efficiency in posterior inference.
- Developed and validated a Multivariate Bayesian Shrinkage Prior (Mt-MBSP) model supporting mixed-type outcomes (continuous, binary, count) with Gibbs sampling.
- Conducted extensive simulation studies comparing Bayesian CRD, Bayesian CRD with imputation, and OLS, evaluating predictive accuracy and convergence via Gelman-Rubin diagnostics.
- Applied Bayesian modeling frameworks to real and simulated datasets, generating reproducible analyses and visualizations that supported model validation and interpretation.

SKILLS
- **Software:** R Studio, Python, C++, SAS, Git, Linux, Matlab
- **Coding:** Data Visualization, Data Preprocessing, Mechanical Learning, Shinyapps, R Packages