Mark Z. Xu

3rd Year PhD Student
Department of Statistics
George Mason University

+1-608-886-2403
zxu27@gmu.edu
My GitHub
My LinkedIn

Education

- **PhD in Statistical Science, George Mason University (GMU)** *09/2021 - present*
  Department of Statistics GPA: 4.0/4.0
- **MS in Data Science, University of Wisconsin Madison (UW-Madison)** *09/2019 - 06/2021*
  Department of Statistics GPA: 3.64/4.0
- **Bachelor in Statistics, University of Science and Technology of China (USTC)** *09/2016 - 06/2021*
  School of the Gifted Young, Department of Statistics and Finance GPA: 3.3/4.3

Work Experience

- **Graduate Teaching Assistant** *09/2021-05/2023*
  Department of Statistics, GMU Fairfax, VA
  - TA for STAT 742 (Optimization for Statistical Modeling), STAT 662 (Multivariate Analysis), STAT 658 (Time Series Analysis and Forecasting), STAT 652 (Statistical Inference), STAT 634 (Case Studies in Data Analysis), STAT 544 (Applied Probability) and STAT 250 (Introductory Statistics I). My responsibilities include holding office hours, responding questions, writing solutions, and grading homework assignments.
- **Undergraduate Lab Assistant** *03/2019-06/2019*
  Department of Statistics and Finance, USTC Hefei, Anhui, China
  - Trained under Professor Yugang Yu, the vice chair of the School of Management, USTC, and his assistant Xiaofei Hu with the project "Application of the shortest path problem in the field of logistics distribution," using the software FlexSim.
- **Summer Intern** *07/2018-09/2018*
  National Bureau of Statistics of China, Zhejiang Branch Quzhou, Zhejiang, China
  - Helped Division Director Xiaoyan Zhou in basic statistical analysis using EXCEL and participated in compiling the *Statistical Yearbook of Quzhou (2018).*

Research Experience

- **A Restricted Penalized Test for Detecting Monotone Bias in a Semi-parametric Model** *02/2023 - present*
  Supervised by Professor Jiayang Sun (GMU).
  - Tools & technologies used: R, High-Performance Computing (HPC).
  - Project description: We are developing a restricted penalized test designed to detect monotone selection bias. The test is derived with the knowledge of isotonic regression; the proof of asymptotic distribution relies on the utilization of empirical processes; and the simulation of power curves is established with our school HPC. A poster award was given regarding part of the work in this paper at the Michael Woodroofe Memorial Conference at the University of Michigan (UMich) this September.
- **Data-driven Approaches to Develop Sleep Phenotypes in Spinal Cord Injury** *07/2023 - present*
  Collaborated project led by Professor Letitia Graves (UTMB) and advisor Jiayang Sun (GMU).
  - Tools & technologies used: RMarkdown.
  - Project description: I am involved in the study of categorizing sleep phenotypes by grouping together risk factors to enable early identification and targeted personalized treatment of sleep disturbances in patients with SCI. My work includes exploratory data analysis, similarity coefficients comparison, correspondence analysis, and ggplot visualization.
- **Analytics for Crowdsourcing: Application to Epidemiology** *06/2022 - 01/2022*
  Supervised by Professor Jiayang Sun (GMU).
  - Tools & technologies used: R.
  - Project description: This paper studied the trend of using crowdsourcing in epidemiology articles and gave recommendations regarding sampling strategy. My work included exploratory data analysis and browsing related articles as proper references.

Class Projects

- **Introduction to Tensor Regression: An Example of Simple Least Square Regression** *09/2022 - 12/2022*
  Class project of STAT 676 (Linear Models and Advanced Regression Models) at GMU.
  - Tools & technologies used: R.
  - Project description: We introduced the simple least square tensor regression algorithm with detailed derivation and simulation results. GitHub link: *https://github.com/Lofia/Project_Tensor_Regression*.

- **Polyphonic Music Generation with Recurrent Neural Network** *02/2021 - 05/2021*
  Class project of STAT 453 (Introduction to Deep Learning and Generative Models) at UW-Madison.
  - Tools & technologies used: Python.
  - Project description: We used an LSTM-based neural network to deal with the polyphonic music generation task. We generated three voice parts for one given theme or melody with the self-designed encoder and decoder. Johann Sebastian Bach's chorales were selected to be our background data. One-hot encoding, data padding and normalization process were involved in our project. GitHub link: *https://github.com/Lofia/Project_453_final*.

- **Yelp Analysis** *09/2020 - 12/2020*
  Class project of STAT 628 (Data Science Practicum) at UW-Madison.
  - Tools & technologies used: R, Python, Github Page App.
  - Project description: By analyzing store information, user information and review comments from Yelp, we aimed to visualize the features of each category of store and give suggestions to restaurant owners for improving their average ratings on Yelp. The data we used are real data from Yelp and we built a Github Page App for users. GitHub link: *https://github.com/AustinJiangH/YelpAnalysis*.

- **Bodyfat Estimation** *09/2020 - 12/2020*
  Class project of STAT 628 (Data Science Practicum) at UW-Madison.
  - Tools & technologies used: R, R Shiny App.
  - Project description: We used linear regression methods to develop an R Shiny App to estimate body fat percentage for any individual and visualized the models using ggPlot. The analysis was based on a real data set of 252 men with measurements of their percentage of body fat and various body circumference measurements. GitHub link: *https://github.com/zli995/Bodyfat-Estimation*.

- **Technical Analysis Price Patterns** *09/2020 - 12/2020*
  Class project of STAT 605 (Data Science Computing Project) at UW-Madison.
  - Tools & technologies used: R, HPC.
  - Project description: We used a large stock dataset from Kaggle to verify if the Technical Analysis of Price Patterns method works in real stock dataset. The work consists of effective and comprehensive data cleaning and large-scale distributed computing. GitHub link: *https://github.com/Lofia/605_final*.

Technical Skills and Interests

Developer Tools: Fluent in R, have been using Python, SQL, MATLAB and C in class projects.
Areas of Interest: Performing statistical data analysis, designing statistical methods, and doing theoretical studies.

Awards

- **Outstanding Poster Presentation** *09/2023*
  Outstanding Poster Presentation Award in the Michael Woodroofe Memorial Conference at UMich.

- **Excellence Scholarship** *2016, 2017, 2018*
  Annual Excellence Scholarship Award at USTC.