Yiren Hou

Ann Arbor, Michigan | 858-952-3525 | yirenhou@umich.edu

EDUCATION

Aug 2023 –	University of Michigan , Ann Arbor, MI
Apr 2025	M.S. in Biostatistics
Aug 2019 – May 2023	University of Georgia, Morehead Honors College, Athens GA B.S. in Statistics, Math Minor, Computer Science Minor Certificate of Applied Data Science Summa Cum Laude, with Highest Honors Undergraduate Thesis: "Some Contributions to Design of Experiments and Survey Sampling"

RESEARCH EXPERIENCE

May 2024 – Jan 2025	UM Department of Biostatistics Supervised by: Dr. Bhramar Mukherjee Submitted manuscript: "Global Prevalence of Long COVID, its Subtypes and Risk factors: An Updated Systematic Review and Meta-analysis" https://doi.org/10.1101/2025.01.01.24319384
	Leading an updated systematic review and meta-analysis on long COVID and its subtypes and symptoms. Responsible for communicating with librarians and team members, working on all the components of systematic review, performing formal quantitative study, and writing the manuscript.
	 Organized substantial background research for long COVID subtypes and extensive search strategies for finding publications Performed data synthesis, risk assessments, and meta-analysis on prevalence and risk factors of long COVID and its subtypes in R Compared traditional systematic review method to GPT applications to systematic review through prompt learning and automatic tools
Sep 2021 – Dec 2023	UGA Department of Statistics Supervised by: Dr. Gauri Datta and Dr. Abhyuday Mandal Submitted manuscript: "Credible Distributions of Overall Ranking of Entities" https://arxiv.org/abs/2401.01833
	Motivated by the joint confidence region constructed from the frequentist approach by Klein, Wright, and Wieczorek (2020), a joint optimal credible region for overall ranking of populations is constructed and is more informative. Based on the Bayesian approach to inference, estimation uncertainty is provided for estimated ranks in small area estimation.
	 Computed in R to apply Bayesian method to overall ranking of states on mean travel time to work in 2011 collected by the US Census Bureau Provided the probabilities of various rankings a state may have based on its mean travel time to work in 2011 and the probabilities that various states may occupy a particular ranking based on their means

CURO Summer Fellowship Project: "Mixture model for Small A rea Estimation using Bayesian Methods"

Proposed a Bayesian approach to small area estimation: two-component mixture model with one area-level random effect from the Normal distribution and another from Cauchy distribution. This model would be robust against outliers and random effects from *t*-distribution with low degree of freedoms.

- Studied the derivation of the proposed model and coded the simulations and Gibbs Samplings in R
- Compared proposed model with other models by coverage probabilities, empirical bias, posterior variance, and interval scores
- Utilized computing clusters for simulations with Linux commands

Sep 2021 -UGA Department of Statistics and Department of Textiles, Merchandising, andApr 2022Interiors

Supervised by: Dr. Abhyuday Mandal and Dr. Suraj Sharma CURO Research Project: "From Rotten Milk to Dressing Gown – Optimal Design for Ordinal Categorical Regression"

Aimed to find the optimal process for creating milk fiber using milk casein protein and other additives. This was an independent research in designing the experiment, running trials, and analyzing data using ordinal analysis.

- Reviewed literature, selected appropriate additives based on their properties, and planned the procedure
- Designed experiment using orthogonal arrays and maximin design criterion from Morris and Mitchell, 1995
- Conducted two hours lab works at least three times a week

ACADEMIC PROJECTS

Feb 2024 –

Apr 2024

Machine Learning Project

Instructor: Dr. Xiang Zhou "Predicting Cancer Patient Survival Time using Random Forests on Spatial

Proteomics Data"

Based on sub-cellular resolution spatial proteomics data, feature extractions and random forest were applied to predict cancer patient survival time in months. Code was developed in Python.

- Conducted non-spatial and spatial-based feature extraction via pixel connectivity, local binary patterns, and histogram of oriented gradients
- Built random forest model by training set and evaluated on testing set
- Identified important features such perimeter, area, and patterns of specific protein-biomarker for prediction

Oct 2023 – Statistical Computing Project

Dec 2023 Instructor: Dr. Hyun Min Kang "SpatialPCA +: Extension of SpatialPCA for Spatial Transcriptomics"

Motivated by dimensionality reduction method SpatialPCA in Shang and Zhou's paper (2022), nearest neighbor Gaussian process (NNGP) was applied

	to approximate inverse of kernel matrix and estimate the low-dimensional embedding of gene expression matrix from the latent factor model in SpatialPCA.
	 Took initiative in method development and integration with SpatialPCA Developed code in R, tested computational efficiency between proposed NNGP method and eigen-decomposition used in SpatialPCA, performed real data analysis, and created an R package Found NNGP was more efficient in approximating inverse of large sparse kernel matrices compared to eigen-decomposition
Aug 2021 – May 2022	Capstone Project Supervised by: Dr. Lynne Seymour and Greg Ellison <i>"Projecting UGA Donations"</i>
	This group project was given by the UGA Development and Alumni Relations. Through an extensive report, the goal was to gauge details on UGA donations from 2013 to 2021 based on demographic variables and make predictions.
	 Led comprehensive exploratory data analysis using base R and tidyverse and supervised time series analysis Reviewed and edited substantial report writings Oversaw schedules for presentation practices and contacted client for questions and updates
Feb 2022 – May 2022	Projects for MATH 4740 Optimization and Data Analysis Instructor: Dr. Malcolm R. Adams <i>"Estimating Total Damage Cost from Tornado Data"</i> <i>"Fourier Analysis for Time Series using Tornado Data"</i>
	For the first project, an estimation study of total damage was done through multiple linear regression. For the second project, Discrete Fourier Transform was used to investigate seasonality. Code was implemented in Python. Tornado data was obtained from NOAA.

PRESENTATIONS

Apr 2023	Overall Ranking of Populations using Bayesian Method. Poster at Center for Undergraduate Research (CURO) Symposium, Athens, Georgia
Oct 2022	Overall Ranking of Populations using Bayesian Method. Poster at Georgia Statistics Day, Athens, Georgia
Oct 2022	Optimal Design for Ordinal Categorical Regression on Milk Fiber Strength. Oral presentation at International Conference on Advances in Interdisciplinary Statistics and Combinatorics (AISC 2022), Greensboro, North Carolina
Jul 2022	Mixture model for Small Area Estimation using Bayesian Methods. Poster at CURO Summer Fellowship Final Forum, Athens, Georgia
Apr 2022	From Rotten Milk to Dressing Gown – Optimal Design for Ordinal Categorical Regression. Poster at CURO Symposium, Athens, Georgia

SCHOLARSHIPS AND AWARDS

Fall 2022	Georgia Statistics Day Undergraduate Poster Presentation 1st Place – Awarded in competition sponsored by the Georgia Statistics Day Conference	
Fall 2022	AISC 2022 Oral Presentation 3 rd Place – 1 of 3 Undergraduate Research Award recipients chosen from top 5 presenters, sponsored by the International Conference on Advances in Interdisciplinary Statistics and Combinatorics	
Fall 2022	UGA Center for Undergraduate Research Opportunities (CURO) Conference Participation Grant – Awarded funding to present research at conferences throughout the United States. Funded for AISC 2022 and advised by Dr. Mandal	
Summer 2022	UGA CURO Summer Research Fellowship – 1 of 30 students selected to receive funding for full-time, intensive faculty-mentored undergraduate research during summer. Mentored by Dr. Datta and Dr. Mandal	
Fall 2021	UGA CURO Research Assistantship – Selected to receive scholarship that supports outstanding undergraduate students to actively participate in faculty-mentored research. Mentored by Dr. Mandal and Dr. Sharma	
Spring 2020	Undergraduate Statistics Class Project Competition (Intermediate) 3 rd Place – Awarded in competition sponsored by American Statistical Association and the Consortium for Advancement of Undergraduate Statistics Education	
Fall 2019 – May 2023	Frank and Louise Hoffmaster Etchberger Scholarship – Awarded renewable scholarship for students with good standing in the Honors College and majoring or minoring in math with a cumulative GPA of 3.6 or higher	

TECHNICAL SKILLS

Proficient	R, Python
Intermediate	Java, LaTeX

WORK EXPERIENCE AND COMMUNITY INVOLVEMENT

Aug 2024 – Dec 2024	 UM LSA Complex Systems 445 Grader Assessed students' assignments with fairness and transparency and managed grading tasks efficiently to meet deadline for grade release
Jul 2024 – Apr 2025	 UM Biostatistics Peer Mentoring Guide incoming graduate students in the department through sharing academic advice and helping the mentees to succeed
Jan 2024 – Present	 Dr. Bhramar Mukherjee Research Lab Attend bi-weekly meetings to seek knowledge in data equity, analysis of electronic health records, and integration of heterogeneous data sources