Yiying (Emma) Wu
Website: https://wuyiying2018.github.io/ Phone: (702)637-5693 E-mail: Yiying.Wu@uth.tmc.edu

EDUCATION

The University of Texas Health Science Center at Houston, School of Public Health Houston, TX
Doctor of Philosophy in Biostatistics 08/2024 – Present

- Passed Preliminary Exam (08/2025)
- Relevant Coursework: Survival Analysis, Statistical Computing, Statistical Inference, Multivariate Analysis, Stochastic Processes

Columbia University Mailman School of Public Health New York, NY
Master of Public Health in Biostatistics 09/2022 - 05/2024

- Cumulative GPA: 4.0 /4.0
- Relevant Coursework: Applied Regression, Data Science, Analysis of Health Survey Data, Database (MySQL)
- Honor:
  Finalist for DataFest 2024 (Top 4, Oral presentation at the ENAR Spring Meeting in Baltimore, Maryland, 03/2024), Delta Omega Honorary Society in Public Health (07/2024)

Shanghai University of Finance and Economics Shanghai, China
Bachelor of Science in Statistics 09/2018 - 07/2022

- Cumulative GPA: 3.1 /4.0
- Relevant Coursework: Real Analysis, Time Series, Survival Analysis, Machine Learning, Stochastic Processes
- Honor: First Prize for the 5th Mathematical Modeling Competition of SUFE (Top 1%, 04/2019)
- Certification: SUFE-Cambridge Leadership Online Programme (07/2020)

RESEARCH

Graduate Research Assistant Houston, TX, MD Anderson Cancer Center (10/2024 – Present)
Project: PCA-based integration algorithm review, NMF model proposal and scHi-C + scRNA-seq integration

- Developed and evaluated statistical methods for integrating scHi-C and scRNA-seq datasets from mouse brain.
- Implemented supervised PCA, reciprocal PCA, and non-negative matrix factorization (NMF) for multimodal integration.
- Applied Seurat pipelines and scGAD transformation for single-cell 3D chromatin data preprocessing.
- Future work includes combining NMF and anchor mapping for improved single-cell multi-omics integration.

Graduate Research Project Houston, TX, 04/2025-05/2025
Project: BMI, Hypertension, and Cardiovascular Biomarkers: A Multivariate Analysis of the Framingham Study

- Applied MANOVA, Hotelling's T^2, and MANCOVA to Framingham Heart Study data (n = 3,828).
- Identified significant associations of BMI and hypertension with cholesterol, glucose, and heart rate.
- Demonstrated independent and additive effects of obesity and hypertension on cardiovascular risk profiles.

Research Assistant New York, NY, 06/2023 - 05/2024
Project: Predicting Methylation from Sequence and Gene Expression Using Deep Learning Methods

- Engineered advanced deep learning algorithms capable of precisely forecasting DNA methylation patterns, integrating genomic sequences with gene expression data for robust biological predictions.
- Employed Python programming within Columbia University's high-performance computing environment to validate and enhance the predictive accuracy and dependability of deep learning models for genomic analysis.
- Achieved a Spearman correlation of 0.84 in model predictions, indicating high accuracy in DNA methylation level prediction.

Graduate Researcher New York, NY, 12/2023-01/2024

Project:Trends in Hypertension Control and Management Disparities in US Adults: A NHANES Analysis from 1999-2020

- Utilized National Health and Nutrition Examination Survey (NHANES) data spanning 1999-2020 to investigate trends in stage 2 hypertension management among U.S. adults.
- Examined prevalence, awareness, and medication use related to hypertension, incorporating demographic factors and comorbidities as covariates in logistic regression models.
- Employed multiple imputation techniques to address missing data and adjusted for survey cycle weights.

Graduate Researcher New York, NY, 10/2023 - 12/2023

Project: Analysis of Homelessness Crisis in California

- Analyzed two key datasets: the first capturing hospital encounters of California's homeless from 2019-2020, and the second providing demographic insights into those accessing homeless services from 2017-2023, to decipher trends in homelessness and healthcare utilization.
- Utilized Exploratory Data Analysis (EDA) and a Shiny App to identify demographic trends in California's homelessness services
- Applied the Yeo-Johnson transformation to linear regression data to enhance normality

Research Assistant in Chinese Academy of Sciences Shanghai, China 06/2021 – 08/2021

Project: A Study of the Existing Machine Learning Models on Reading Comprehension

- Conduct a literature review of current NLP methods
- Leveraged external knowledge-enhanced models to speedily eliminate potential scenarios that are inconsistent with common sense, thereby dramatically reducing computational effort and improving efficiency
- Proposed an interpretable three-module multi-hop reading comprehension system, EPAr, which constructed an inference tree, proposed a single response candidate for each root-to-leaf chain, and consolidated key information from all inference chains for the prediction

Independent Team Leader Shanghai, China 10/2020 – 12/2020

Project: A Study on the Healthcare Service of Residents in Chongming District, Shanghai

- Led a team of 3 that conducted a field survey of 120 households to study the healthcare service of local residents using STATA to perform descriptive analysis and statistical inference
- Mapped the relationship between healthcare coverage, the size of the community, and the average age of the community
- analyzed the status and deficiencies of rural healthcare, and provided suggestions and solutions for policymakers to improve rural healthcare based on the analysis results

INTERNSHIPS

Data Analyst *Wellness Equity Alliance* Vernon, CA, 07/2023 – 09/2023

- Conducted comprehensive demographic data analysis employing Stata and Excel to inform public health initiatives
- Assisted in the development of Vernon, California's public health infrastructure, enhancing service delivery for the local and surrounding communities
- Leveraged Google Earth (GIS) for the effective visualization of demographic trends, the distribution of mental health facilities, and the availability of food security resources, facilitating improved community access and resource planning

AI Analyst *Microsoft China, Microsoft AI and IoT Insider Lab* Shanghai, China, 06/2021 - 08/2021

- Tested the Microsoft Azure cloud product "deep-voice-conversion-master" through converting audio text data into target text with Machine Learning, and wrote code in Python to convert the target text into the specified format in batch
- Researched the principles and applications of keypoint detection for human body projects by employing the human key point recognition from open source on GitHub, as well as executing the program in a Linux environment

to achieve the recognition of video and picture

- Employed convolutional neural networks for training in the way of deriving the association and orientation of all the joints on the human bodies in the diagram, identifying and labeling each joint, and outputting the key anatomical points of each person through visualization

Investment Analyst *Hwabao Securities* Shanghai, China, 06/2020 - 08/2020

- Performed empirical analysis of the trend of the pharmaceutical industry using SQL and Excel with 20,000+ data crawled from the internet, and produced quantitative analysis reports
- Examined historical and current market supply and demand of the pharmaceutical industry, and conducted statistical description and regression analysis to predict the risk of overseas investment and surplus capacity
- Engaged in research on the stock of Fuxiang Pharmaceutical, and analyzed the company's historical financial reports using STATA to identify the company's investment potential and core market competencies

SKILLS

- Software: R, MySQL, Python, SAS, Stata, LaTeX, MATLAB, and MS Office
- Language: English (Fluent), Chinese (Native)