Chuqi Wang

Irvine, CA 92617 | (949) 849-3189 | chuqiw4@uci.edu

RESEARCH INTERESTS

Statistical methods and applications, statistical modeling and data mining, causal inference, biostatistics, with a focus on health-related problems, including mental health, environmental issues, food safety, public health, and other related areas. I am particularly motivated to explore statistical methods in interdisciplinary research, such as personalized medicine and mobile health.

EDUCATION

University of California, Irvine ***Irvine, CA, USA***
Master of Data Science *Sept. 2023 – Dec. 2024*

- GPA: 3.97/4.0
- Relevant Courses: Databases & Data Management, Probability & Statistical Theory, Statistical Methods, Bayesian Data Analysis

McGill University ***Montreal, QC, Canada***
Bachelor of Science in Statistics *Sept. 2018 – May 2022*

- GPA: 3.56/4.0
- Relevant Courses: Algorithm & Data Structures, Advanced Calculus, Algebra & Analysis, Statistical Learning, Mathematical Statistics, Generalized Linear Models

PUBLICATION

- Wang, Chuqi. "A review on 3d convolutional neural network." 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2023.
- Zhang, Liwen, et al. "Fineval: A chinese financial domain knowledge evaluation benchmark for large language models." arXiv preprint arXiv:2308.09975 (2023).

EXPERIENCE

PFAS Contamination Data Analysis Project *Jun. 2024 – Feb. 2025*

- Collected and processed 2.5 million PFAS-related water quality records from public data sources across all 50 U.S. states.
- Conducted spatial and statistical analysis using Python to identify regional trends in PFAS contamination.
- Reviewed academic literature and applied statistical learning methods to support predictive research and contribute to a collaborative publication.

Language Model Evaluation Research *Jun. 2024 – Oct. 2024*

- Conducted academic literature reviews and collected domain-specific datasets to support a survey study on large language models (LLMs).
- Evaluated the performance of 19 LLMs across various reasoning and comprehension tasks using a structured benchmark.
- Co-authored a paper on the FinEval benchmark, contributing data presentation through extensive tables and figures formatted in LaTeX.

PROJECTS

Stroke Prediction Using Bayesian Logistic Regression *Feb. 2024 – Mar. 2024*

- Developed a Bayesian logistic regression model using the rstan package in R, fitted with 2000 iterations and 4 Markov chains via Markov Chain Monte Carlo (MCMC), to predict stroke occurrence based on 5110 patients' demographic, medical, and lifestyle data. Achieved a 95.2% test accuracy and improved model sensitivity through decision threshold adjustment.
- Performed data preprocessing and exploratory data analysis (EDA) using dplyr and ggplot, and applied diagnostic tools like Bulk ESS, Tail ESS, and trace plots to ensure model convergence and reliability.

SKILLS

Programming Languages: Python, R, Java, SQL

Frameworks & Tools: Jupyter, SciKit-Learn, TensorFlow, Tableau, LaTeX

Statistical Methods: Bayesian inference, Logistic regression, Data visualization