ZHENYU XU

959-929-9579 | zhenyu.xu@uconn.edu

EDUCATION

Ph.D. in Statistics, University of Connecticut, Storrs, CT (GPA: 3.91/4.0) 08/2021 - Expected 2025

MSc in Data Science, University College London, UK (Distinction) 09/2019 - 09/2020

BSc in Actuarial Science, Heriot-Watt University (First-Class Honors, GPA: 4.0/4.0) 09/2017 - 05/2019
Awards: Deputy Principal's Award (07/2018 and 07/2019)

BSc in Actuarial Science, Southwestern University of Finance and Economics, China 09/2015 - 07/2019
Awards: Third-Class Scholarship (12/2017)

TECHNICAL SKILLS

- **Programming Language**: Python, SQL, R, SAS, Git, Bash
- **Data Science**: Machine Learning (scikit-learn, XGBoost), Deep Learning (TensorFlow), Data Visualization (Tableau, ggplot2), A/B Testing, Hive
- **Statistics**: Probability Theory, Mathematical Statistics, Generalized Linear Model, Time Series Analysis, Survival Analysis, Statistical Computing, Analysis of Variance, Bayesian Statistics, Statistical Learning, Experiment Design

PROFESSIONAL EXPERIENCE

UConn Statistical Consulting Services Storrs, CT
Statistical Consultant 05/2023 - 08/2023

- Tutored a group of 30 students from diverse background on data manipulation and visualization using R in a workshop
- Created boxplots of scores for pre and post survey data in a chemistry lab using ggplot2, and examined the difference between pre and post test scores based on the question type using t-test, Cohen's D and ANOVA
- Identified the responsive minor-like introns for different treatments and reduced the type II errors

DataGrasp | Medicspot Team London, UK
Statistical Consultant 10/2019 – 12/2019

- Processed Medicspot's data (pharmacy addresses and installed device) to construct a regression model
- Calculated and ranked the expected revenue of installing Medicspot device in pharmacies across the UK

RESEARCH

Child Maltreatment Identification, UConn Health 05/2023 - Present

- Combined the encounter, diagnosis and demographic datasets and selected the maltreatment group by their ICD-10 codes
- Generated contingency tables for each predictor in relation to child maltreatment occurrence and determined potential risk factors through the application of both the Chi-square test and Fisher's exact test.
- Adjusted the p values to control the false discovery rate
- Implemented a logistic regression classifier to identify child maltreatment

EnergyStats, UConn 05/2023 - Present

- Maintained and Developed the EnergyStats website using Node.js, Pug, and CSS
- Utilized the bi-clustering method to simultaneously cluster building accounts and respective months based on their energy usage data on campus

On GEE for Mean-Variance-Correlation Models: Variance Estimation and Model Selection, UConn 05/2022 - Present

- Specified the correct joint model for mean-variance-correlation structure using Generalized Estimating Equations
- Identified the correct sandwich variance estimator for parameter estimates
- Conducted simulation studies to investigate the differences between the two models and variance estimators
- Developed a new criterion for joint model selection to achieve a correct selection rate of **95.6%**

(Master's Dissertation) NIR Spectroscopic Data Classification by Ensemble Methods, UCL 06/2020 - 09/2020

- Used PCA to reduce dimensions to **7** and maintained **95%** information of high-dimensional spectral data
- Employed two ensemble methods (Random Forest and XGBoosting) to classify the data
- Used a double cross-validation scheme to evaluate each model's performance

Information Retrieval Model Project, Information Retrieval and Data Mining, UCL 01/2020 - 04/2020

- Built a passage re-ranking system of a given candidate list of passage to a query using retrieval models
- Built an inverted-index for the collection for efficient passage retrieval, tried retrieval models such as BM25 model, logistic regression, LambdaMART Model and neural network, with Word2Vec embedding

ACADEMIC PRESENTATIONS

- Variance estimation for generalized estimation equations of mean-variance-correlation for clustered data, **Excellence in Statistical Science**, Storrs, CT, 2022
- Disparity in county-level low-income job loss rate during the Covid-19 pandemic, **The Joint Statistical Meetings**, Washington, DC, 2022