

## EDUCATION

---

<b>Ph.D. in Statistics</b> , University of Connecticut, Storrs, CT (GPA: 3.9/4.0) <i>Advisor:</i> Dr. Jun Yan, Professor, Department of Statistics, University of Connecticut	08/2021 - Expected 2025
<b>MSc in Data Science</b> , University College London, UK (Distinction)	09/2019 - 09/2020
<b>BSc in Actuarial Science</b> , Heriot-Watt University, UK (First-Class Honors, GPA: 4.0/4.0) <i>Awards:</i> Deputy Principal's Award (07/2018 and 07/2019)	09/2017 - 05/2019
<b>BSc in Actuarial Science</b> , Southwestern University of Finance and Economics, China <i>Awards:</i> Third-Class Scholarship (12/2017)	09/2015 - 07/2019

## TECHNICAL SKILLS

---

- **Programming Language:** Python, SQL, R, SAS, Git, Bash
- **Data Science:** Machine Learning (scikit-learn, XGBoost), 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

---

<b>UConn Statistical Consulting Services</b> Statistical Consultant	Storrs, CT 05/2023 - 08/2023
<ul style="list-style-type: none"><li>• Tutored a group of 30 students from diverse background on data manipulation and visualization using R in a workshop</li><li>• 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</li><li>• Identified the responsive minor-like introns for different treatments and reduced the type II errors</li></ul>	
<b>DataGrasp   Medicspot Team</b> Statistical Consultant	London, UK 10/2019 - 12/2019
<ul style="list-style-type: none"><li>• Processed Medicspot's data (pharmacy addresses and installed device) to construct a regression model</li><li>• Calculated and ranked the expected revenue of installing Medicspot device in pharmacies across the UK</li></ul>	

## RESEARCH

---

<b>(Ph.D. Dissertation) Generalized Estimating Equations, UConn</b>	05/2022 - Present
<ul style="list-style-type: none"><li>• Specified the correct joint model for mean-variance-correlation structure using Generalized Estimating Equations</li><li>• Identified the correct sandwich variance estimator for parameter estimates</li><li>• Conducted simulation studies to investigate the differences between the two models and variance estimators</li><li>• Developed a new criterion for joint model selection to achieve a correct selection rate of <b>95.6%</b></li></ul>	
<b>Child Maltreatment Identification, UConn Health</b>	05/2023 - Present
<ul style="list-style-type: none"><li>• Combined the encounter, diagnosis and demographic datasets and selected the maltreatment group by their ICD-10 codes</li><li>• 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.</li><li>• Adjusted the p values to control the false discovery rate</li><li>• Implemented a logistic regression classifier to identify child maltreatment</li></ul>	
<b>EnergyStats, UConn</b>	05/2023 - Present
<ul style="list-style-type: none"><li>• Maintained and Developed the EnergyStats website using Node.js, Pug, and CSS</li><li>• Utilized the bi-clustering method to simultaneously cluster building accounts and respective months based on their energy usage data on campus</li></ul>	
<b>(Master's Thesis) NIR Spectroscopic Data Classification by Ensemble Methods, UCL</b>	06/2020 - 09/2020
<ul style="list-style-type: none"><li>• Used PCA to reduce dimensions to <b>7</b> and maintained <b>95%</b> information of high-dimensional spectral data</li><li>• Employed two ensemble methods (Random Forest and XGBoosting) to classify the data</li><li>• Used a double cross-validation scheme to evaluate each model's performance</li></ul>	
<b>Information Retrieval Model Project, Information Retrieval and Data Mining, UCL</b>	01/2020 - 04/2020
<ul style="list-style-type: none"><li>• Built a passage re-ranking system of a given candidate list of passage to a query using retrieval models</li><li>• 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</li></ul>	

## 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