



## Zefang Min

Department of Statistics  
University of Connecticut  
215 Glenbrook Road, Unit 4120  
Storrs, CT 06269–4120 USA

(860)931-8799  
✉ [zefang.min@uconn.edu](mailto:zefang.min@uconn.edu)

[GitHub Profile](#)  
[Personal Website](#)

### RESEARCH INTEREST

---

Time Series, Causal Inference, Neural Networks

### EDUCATION

---

- **University of Connecticut, Storrs** 2023 - Present  
*Ph.D. in Statistics (Advised by Jun Yan)*
- **University of Connecticut, Storrs** 2021 - 2023  
*M.S. in Statistics*
- **Central University of Finance and Economics, Beijing, China** 2016 - 2020  
*B.A. in Public Finance and Public Policy*

### EXPERIENCE

---

- **PwC(China)** Jan 2019 - Feb 2019  
*Tax Winter Intern* Beijing
  - Researched Chinese tax laws and administrative documents.
  - Assisting my colleagues in providing tax dispute resolutions.

### PERSONAL PROJECTS

---

- **Data Scraping for Gymnastics Data** July 2023  
*As a preparation for The 5th UConn Sports Analytics Symposium in 2024*
  - \* Wrote R functions that extract data from PDF tables and transform them into the format suitable for analysis.
  - \* Improved extract—table function from Tabulizer by writing a function automatically getting page areas.
  - \* Scraped data from more than 20 gymnastic championships, cups, and tournaments from PDFs from 2017 to 2023.
- **2021 Travelers Model Competition** Oct 2021 - Nov 2021  
*To create a predictive model for fraud detection based on historical claim data, launched through Kaggle*
  - \* Cooperated with two teammates, and won the campus winner in UConn. The score of our model ranked first in all 22 teams that competed (final private leaderboard).
  - \* Constructed an ensemble boosting model consisting of XGBoost, LightGBM, and AdaBoost using Python, did a thorough data preprocessing, derived some new features from the data, and analyzed the feature importance using LIME.
  - \* Gave the final presentation for our team on the Job Shadow Day.
- **A Phase II Clinical Research Project for Statistics in Pharmaceuticals Summer Course** July 2022  
*Focus on PASI score data (ADPA dataset), simulate missing data, and analyze data using different approaches*
  - \* Built up logistic regression model to investigate the relation between PASI75 and treatment and sex under different missing patterns, using R and SAS
  - \* Investigated the bias of missing value handling methods under different missing patterns.

### TECHNICAL SKILLS

---

**computing skills:** R, Python, Git, Latex, Html, Julia(learning), SAS(learning)  
**Language:** Chinese(native), English(proficient)

### RELATED COURSES TAKEN

---

**Graduate-level Courses:** Mathematical Statistics I II, Applied Statistics I II, Linear Model, Design of Experiments, Intro to Data Science, Applied Time Series, Intro to Biostat  
**Undergraduate-level Courses:** Statistical Computing, Non-Parametric Methods, Elementary Stochastic Process, Econometrics