

Qingkai Dong

Email: qingkai.dong@uconn.edu Website: dqksnow.github.io
Phone: +1 (860) 931-9075 Location: Storrs, CT, USA

Education

University of Connecticut Ph.D. in Statistics Advisors: Prof. HaiYing Wang and Prof. Jun Yan	Aug 2023 – Expected 2027 Storrs, CT, USA Passed the Ph.D. Qualifying Exam
Zhongnan University of Economics and Law MS in Mathematical Statistics Advisor: Prof. Hui Zhao Thesis: Model Averaging and Variable Selection for Accelerated Failure Time Models	Sep 2020 – May 2023 Wuhan, China
Qingdao University BS in Applied Statistics Thesis: Automatic Marking System Based on Convolutional Neural Networks	Sep 2016 – May 2020 Qingdao, China

Research Experience

Global Biometrics Intern Servier Pharmaceuticals	Jun 2026 – Aug 2026
- Designed and implemented a modular workflow to reconstruct synthetic individual participant data from published Kaplan–Meier curves and summary statistics using target-matching loss functions. - Validated close reproduction of reported survival metrics and recovery of unreported quantities through simulation studies; applied the workflow to internal data.	
Research Assistant (Academia–Industry Collaboration) Servier Pharmaceuticals & University of Connecticut	Mar 2024 – Dec 2025
- Coauthored a manuscript introducing a predictive-modeling–assisted interim analysis framework for censored time-to-event endpoints; manuscript under revision. - Developed a covariate-informed approach that augments interim data by predicting event times for censored participants, improving estimation of conditional power and interim decision-making. Implemented in PyTorch. - Proposed evaluation metrics for conditional power forecast accuracy and futility decision quality; conducted extensive simulations to characterize when performance improves or degrades under varied censoring, covariate informativeness/type and prediction range.	
Research Assistant (Statistical Methodology) University of Connecticut	Aug 2023 – Present
- Coauthored a methodology paper on rare-feature-aware subsampling for regression models; manuscript in preparation for journal submission. - Developed theory showing that estimation efficiency for rare binary covariate coefficients is driven by the limited number of rare observations, explaining numerical instability and slow convergence. - Introduced balanced subsampling methods that quantifies rarity to ensure adequate representation of rare features without pilot sampling, improving robustness and efficiency. - Implemented the methods in the subsampling R package; validated performance via theoretical results, simulation studies, and real-data applications; produced documentation and reproducible examples.	
Research Assistant (Social Determinants of Health, Frailty, and Accelerated Aging in Breast Cancer Survivors) Jul 2024 – Aug 2024 Departments of Statistics and Human Development and Family Sciences, University of Connecticut	
- Helped frame and prepare a study examining social determinants of health, frailty, and accelerated aging among breast cancer survivors. - Conducted data retrieval, cleaning, and integration across All of Us EHR, survey, and physical-measurement data to support subsequent statistical modeling.	
Cross-Disciplinary Collaboration (Chemical Sensor Data Analytics) Departments of Statistics and Chemistry, University of Connecticut	Mar 2024 – May 2024
- Analyzed fluorescence response data from porphyrinoid sensors; applied clustering and statistical analysis to improve compound differentiation and sensor selection. - Produced interpretable summaries and visualizations in R to support experimental decision-making.	

Cross-Disciplinary Collaboration (Large Language Model Explainability)
University of Connecticut & collaborators

2025 – Present

- Coauthored a survey manuscript on time series explainability emphasizing LLM-enabled semantic explanations; under review at a top survey journal; released an accompanying benchmark repository on GitHub.

Software

R package subsampling: A statistical computing toolkit for optimal subsampling in big-data modeling. Supports GLMs, quantile regression, softmax regression and rare event/rare feature regimes.

- Implemented optimized sampling probability computation targeting variance reduction and prediction accuracy, including practical routines for pilot sampling and two-step designs.
- Engineering focus: modular API, documentation/vignettes, reproducible examples, and scalable computation patterns.

Teaching Experience

Teaching Assistant

Sep 2026 – Dec 2026

University of Connecticut

- **STAT 110Q: Elementary Concepts of Statistics; STAT 3445: Introduction to Mathematical Statistics II.**

Principal Instructor

Jan 2026 – May 2026

University of Connecticut

- **STAT 3675Q: Statistical Computing.** Independently taught three 50-minute sections serving approximately 20 students; designed course slides and R-based homework assignments, coordinated student presentations, and graded midterm and final examinations.
- Delivered a comprehensive R/RStudio curriculum covering data acquisition and management, visualization, reproducible workflows, regression, nonparametric inference, GLMs, categorical data analysis, time series, and classification.

Academic Presentations

Invited Talk - Joint Statistical Meetings (JSM) 2026

Invited Talk - New England Statistics Symposium (NESS) 2026

Invited Talk - ICSA Applied Statistics Symposium 2026

Poster Presentation - New England Statistics Symposium (NESS) 2026

Poster Presentation - The Design and Analysis of Experiments (DAE) 2026

Poster Presentation - New England Rare Disease Statistics (NERDS) Workshop, Boston, MA 2025

Poster Presentation - Dahshu Data Science Symposium, Storrs, CT 2025

Academic Service

Journal Reviewer: *Computational Statistics and Data Analysis*; *Journal of Computational Science*; *Statistical Papers*; *Sankhya B*; *Journal of Systems Science and Mathematical Sciences (Chinese)*

Awards

UConn Graduate School Conference Participation Award Jul 2026

New England Statistics Symposium (NESS) Poster Award 2026

ICSA Applied Statistics Symposium Student Paper Award 2026

Predoc Fellowship - University of Connecticut 2025

First-class Scholarship - Zhongnan University of Economics and Law 2020, 2022

Technical Skills

Statistical Methods: survival analysis; interim analysis & conditional power; design simulation; big-data subsampling; model averaging; variable selection

Programming: R (package development), Python (machine learning / deep learning), Git, LaTeX

Selected Papers

Dong Q, Yan J, Zhang Q, Wang H. Preserving Rare Features in Big Data Regression: Balanced Subsampling. *In preparation for submission.*

Dong Q[†], Man A[†], Hua Z, Teng Z, Wang H, Yan J. Predictive-Modeling-Assisted Interim Decision-Making in Adaptive Trials with Censored Survival Outcomes. *Under revision.* [†]Equal contribution.

Dong Q, Liu B, Zhao H. Weighted Least Squares Model Averaging for Accelerated Failure Time Models. *Computational Statistics and Data Analysis*, 2023.

Zhao H, **Dong Q**. Variable Selection for Additive Hazards Model with Current Status Data. *Journal of Systems Science and Mathematical Sciences*, 2022.

Zhao H, Liu B, **Dong Q**. Jackknife Model Averaging of AFT Model with Current Status Data. *Acta Mathematicae*

Applicatae Sinica, 2023.

Chen Z, Lucchesi G, **Dong Q**, Zheng X, Song D, Wen Q, Cheng W, Ni J, Luo D. From Signals to Semantics: A Survey on Time Series Explainability through a Human-Cognitive Lens. Manuscript under review at *ACM Computing Surveys*, 2025.