Qingkai Dong

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

Education

University of Connecticut Aug 2023 – Present
PhD in Statistics Storrs, CT, USA
Advisors: Prof. HaiYing Wang and Prof. Jun Yan Passed the Ph.D. Qualifying Exam

Zhongnan University of Economics and Law Sep 2020 – May 2023
MS in Mathematical Statistics Wuhan, China
Thesis: Model Averaging and Variable Selection for Accelerated Failure Time Models

Qingdao University Sep 2016 – May 2020
BS in Applied Statistics Qingdao, China

Research Experience

Research Assistant (Academia–Industry Collaboration) Mar 2024 – Dec 2025
Servier Pharmaceuticals & University of Connecticut

Primary output: Coauthored a manuscript introducing a predictive-modeling–assisted interim analysis framework for censored time-to-event endpoints; manuscript in preparation for journal submission.

Developed a covariate-informed approach that augments interim data by predicting event times for censored participants, improving estimation of treatment effects and conditional power.

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, prediction range, and sample size.

Research Assistant (Statistical Methodology) Aug 2023 – Present
University of Connecticut

Primary output: 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 a balanced subsampling method 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 (Real-World Health Data) Jul 2024 – Aug 2024
University of Connecticut

Conducted data preprocessing and statistical analyses on All of Us data to study associations among social determinants of health, frailty index, and accelerated aging in breast cancer patients.

Delivered analysis-ready datasets and reproducible scripts; performed exploratory and confirmatory analyses under missingness and data-quality constraints.

Research Project (Chemical Sensor Data Analytics) Mar 2024 – May 2024
University of Connecticut

Analyzed fluorescence response data from porphyrinoid sensors; applied clustering and statistical analysis to improve compound differentiation and sensor selection.

Produced interpretable summaries and visualizations to support experimental decision-making.

Cross-Disciplinary Collaboration (LLM Explainability) 2025 – Present
University of Connecticut & collaborators

Coauthored a survey manuscript on time series explainability with an emphasis on LLM-enabled semantic explanations; curated benchmarks and an accompanying repository.

Software & Open-Source

R package subsampling (CRAN): A statistical computing toolkit for optimal subsampling in big-data modeling. Supports GLMs, quantile 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

Instructor Aug 2025 – May 2026
University of Connecticut

STAT 2255: Statistical Programming — Introduction to statistical programming via Python, including data types, control flow, object-oriented programming, GUI-driven applications, data wrangling, visualization, and prediction/classification models.

Academic Presentations

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

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

Academic Service

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

Awards

Predoc Fellowship at University of Connecticut 2025

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

Technical Skills

Statistics/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 Publications

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. *Under review*, 2025.