Journal of Data Science 23 (4), 575–577
October 2025
DOI: 10.6339/25-JDS234EDI
Editorial

Editorial: Statistical Frontiers of Data Science

Xiaoling Lu[1,*], Yixuan Qiu[2], Zhezhen Jin[3], Chunming Zhang[4], and Wenxuan Zhong[5]

[1]*School of Statistics, Renmin University of China, China*
[2]*School of Statistics and Data Science, Shanghai University of Finance and Economics, China*
[3]*Department of Biostatistics, Columbia University, USA*
[4]*Department of Statistics, University of Wisconsin–Madison, USA*
[5]*Department of Statistics, University of Georgia, USA*

In this special issue, we present nine articles honoring Prof. Xizhi Wu for his transformative contributions to statistics and data science in China. Over his distinguished career at Nankai University and Renmin University of China, Prof. Wu was a pioneer in integrating modern statistical computing into academic curricula in China. His textbooks emphasize statistical thinking and implementation with open-source tools rather than derivation-heavy theory, reflecting his belief that statistical insight arises through computation and application. Even after retirement, he has continued teaching statistical computing, machine learning, and deep learning across institutions, elevating the visibility of data science within statistical education nationwide. Many of his students and mentees have become international leaders, testifying to his enduring impact and vision that continue to shape the evolution of the field.

Statistical Data Science. The papers under this section showcase methodological innovations that expand the reach of classical models through flexible structures and modern computation. Lu et al. (2025) introduce the smoothing spline state space (QuadS) model that embeds a hidden Markov process with spline-modeled transition and emission matrices, enhancing both predictive accuracy and interpretability. Zhou et al. (2025) extend the Cox model to correlated survival data by replacing the linear risk function with a neural network, combining statistical rigor with the representational power of deep learning. Together, these works exemplify the fusion of traditional modeling and contemporary machine-learning frameworks.

A second group of contributions addresses challenges in high-dimensional data integration and structured inference. Cai et al. (2025) propose EMixed, a unified framework for single- and multi-omics cellular deconvolution that models RNA and DNA methylation data using allocation-based EM estimation. Li et al. (2025) develop gPOCRE, a generalization of penalized orthogonal-components regression for categorical outcomes, constructing orthogonal components that capture shared signal across multiple responses. Both studies advance dimension-reduction and probabilistic modeling in complex biological and high-dimensional settings.

Complementing these developments, Zhang and Ma (2025) focus on comparative medical studies involving correlated bilateral and unilateral data. Their model-based approach treats repeated measurements, such as those from paired organs, within a unified framework and compares performance with existing maximum-likelihood procedures. This work contributes to rigorous inference for specialized experimental designs where correlation structures need to be carefully addressed.

Computing in Data Science. From a computational perspective, Gao et al. (2025) develop an efficient workflow for representing and solving piecewise linear-quadratic optimization

*Corresponding author. Email: xiaolinglu@ruc.edu.cn.

problems via ReLU–ReHU decomposition, implemented in the Python package `plqcom`. Their work exemplifies the growing emphasis on computation and software implementation that translates methodological advances into reproducible tools.

Data Science in Action. Shao (2025) proposes a nonparametric algorithm for multi-step interval forecasting of autoregressive time series using kernel distribution estimation. By avoiding restrictive distributional assumptions, this approach yields flexible prediction intervals that are robust to deviations from normality—an essential feature for reliable forecasting in real-world data applications.

Data Science Reviews. Zhang et al. (2025) provide a selective review of neural-network algorithms for supervised and unsupervised learning, including feedforward, LSTM, and convolutional architectures, with MATLAB implementations to facilitate understanding by statisticians and data scientists. Their tutorial-style presentation bridges the gap between algorithmic development and statistical practice, making advanced methods accessible to a broad audience.

Data Science Conversation. Liu et al. (2025) present a conversation with Prof. Xizhi Wu that traces his academic journey, philosophy of education, and perspectives on the evolution of statistics in the era of artificial intelligence. Through this reflective dialogue, readers gain insight into his lifelong dedication to advancing statistical education and nurturing future generations. The interview provides a multidimensional portrait of an educator whose influence continues to shape the discipline's trajectory.

We extend our sincere gratitude to Editor Dr. Jun Yan for his guidance, the authors for their excellent contributions, and the referees for their constructive reviews. We also thank the reproducibility-checking team at Renmin University of China, led by Dr. Jing Zhou, for ensuring that all code supplements reproduce the reported results. Their efforts reaffirm reproducibility as a hallmark of the *Journal of Data Science*.

References

Cai M, Zhao K, Huang P, Celedon J, McKennan C, Chen W, et al. (2025). *EMixed: Probabilistic multi-omics cellular deconvolution of bulk omics data. Journal of Data Science*, 23(4): 592–606. https://doi.org/10.6339/25-JDS1170

Gao T, Dai B, Qiu Y (2025). ReLU-ReHU representations of piecewise linear-quadratic losses. *Journal of Data Science*, 23(4): 648–658. https://doi.org/10.6339/24-JDS1162

Li Y, Alemdjrodo K, Lin Y, Zhang M, Zhang D (2025). Exploring massive risk factors of categorical outcomes via supervised dimension reduction. *Journal of Data Science*, 23(4): 607–623. https://doi.org/10.6339/25-JDS1188

Liu M, Li J, Wei T, Wang X, Lu X (2025). A journey of wisdom and impact: A conversation with Dr. Xizhi Wu. *Journal of Data Science*, 23(4): 695–715. https://doi.org/10.6339/25-JDS1204

Lu H, Cheng H, Wang Y, Xie Y, Yan H, Wang X, et al. (2025). Mortgage prepayment modeling via a smoothing spline state space model. *Journal of Data Science*, 23(4): 578–591. https://doi.org/10.6339/25-JDS1165

Shao Q (2025). Interval forecasting in time series analysis: Application to COVID-19 and beyond. *Journal of Data Science*, 23(4): 659–675. https://doi.org/10.6339/25-JDS1187

Zhang C, Zhang Z, Zhong X, Li J, Zhao Z (2025). A statistician's selective review of neural network modeling: Algorithms and applications. *Journal of Data Science*, 23(4): 676–694. https://doi.org/10.6339/25-JDS1167

Zhang X, Ma C (2025). Analysis of bilateral and unilateral data: A comparative review of model-based and MLE-based methods for the homogeneity test of proportions. *Journal of Data Science*, 23(4): 638–647. https://doi.org/10.6339/25-JDS1168

Zhou R, He K, Wang D, Liu L, Ma S, Qu A, et al. (2025). Neural network for correlated survival outcomes using frailty model. *Journal of Data Science*, 23(4): 624–637. https://doi.org/10.6339/25-JDS1173