ZEWEI LIN

+1(513) 208-9724 linzw@mail.uc.edu ⋄ zewei-lin.github.io
Carl H. Lindner College of Business, University of Cincinnati
2906 Woodside Drive, Cincinnati OH 45221

EDUCATION

Ph.D. in Business Administration (Business Analytics, GPA: 4.0/4.0),
Carl H. Lindner College of Business, University of Cincinnati 04/2024 (Expected)

B.S. in Mathematical Statistics | B.A. in Philosophy | B.Ec. in Economic Statistics,
Renmin University of China 09/2015 - 06/2019

Summer Session Certificate in Biostatistics,
University of California, Berkeley 07/2017 - 08/2017

RESEARCH

Research Interests:

Business Analytics, Model Transparency, Model Diagnostics, Discrete Data, Network Inference, Statistical Inference/Machine Learning in Insurance/Information System.

——To fulfill business needs and regulatory requirements, my research focuses on developing statistical/machine learning methods to address the multifaceted challenges associated with **model transparency**. The goal is to understand the inner workings of complex models, thereby promoting **more transparent, trustworthy, and interpretable data-driven decision-making models**. I primarily work on problems that involve **discrete data (e.g., binary, rating or count data)** which amplify statistical challenges and call for new developments.

Published Papers:

- Liu, D., Zhu, X., Greenwell B., & **Lin, Z.** (2022), "A new goodness-of-fit measure for probit models: surrogate R^2", *British Journal of Mathematical and Statistical Psychology*, 76, 192-210. https://doi.org/10.1111/bmsp.12289

Working Papers:

- Liu, D., **Lin, Z.**, & Zhang, H. "A unified framework for residual diagnostics in generalized linear models and beyond", submitted to the *Journal of the American Statistical Association*, https://arxiv.org/abs/2207.04299.
- **Lin, Z.**, Liu, D. & Bauer, D., "Unfolding Tweedie regression model for insurance premium pricing: a diagnostic tool leading to actionable insights.", in preparation for *Management Science*.
- Zhu, X., **Lin, Z.**, Liu, D. & Greenwell B., "*Surr_rsq*: an R package for evaluating goodness of fit using surrogate R^2", under review at *the New England Journal of Statistics in Data Science*. [Package]
- **Lin, Z.**, Liu, D. & Samuel, B.,"Joint modeling of multivariate discrete outcomes? An exploratory framework and its application for the design of information system", in progress.

Research in Progress:

- "Bootstrap estimation for sparse edge-exchangeable network," with Yichen Qin.
- "Analyzing Conflicting Information via Multi-dimensional Textual Network Analysis Framework," with Tianhai Zu

Research during college study:

- An online algorithm to calculate high dimensional correlation matrix for analysis of brain image data (Summer 2018, with Moo K. Chung, University of Wisconsin-Madison).
  - It is adopted in *Brain Network Analysis, Cambridge University Press, Chung, M.K. (2019), Page 127.*

AWARDS

- OBAIS Department Graduate Student Teaching Award Recipient 02/2023
- Student and Early-Career Travel Awards, Symposium on Data Science and Statistics 06/2022
- Student Poster Awards (Sponsored by Munich Re), New England Statistics Symposium 05/2022
- *Mingde* Excellent Student Scholarship, Renmin University of China 06/2017

PRESENTATIONS

- **Contributed session**, "*Unfolding Tweedie model for insurance pricing: a diagnostic tool leading to actionable insights*", INFORMS Annual Meeting, Phoenix, AZ. 10/2023(forthcoming)
- **Contributed papers session**, "*Unfolding Tweedie model for insurance pricing: a diagnostic tool leading to actionable insights*", the Joint Statistical Meetings (JSM), Toronto, ON. 08/2023
- **Contributed poster**, "*A unified framework for residual diagnostics in generalized linear models and beyond*", ICSA Applied Statistics Symposium, Ann Arbor, MI. 06/2023
- **Invited session**, "*Surrogate R^2: a new goodness-of-fit measure and an R package for categorical data analysis*", New England Statistics Symposium (NESS), Boston, MA. 06/2023
- **Contributed poster**, "*Unfolding Tweedie model for insurance pricing: a diagnostic tool leading to actionable insights*", The Eighth Bayesian, Fiducial and Frequentist conference (BFF8), Cincinnati, OH. 05/2023
- **Contributed poster**, "*Model diagnostics of discrete data regression: a unifying framework using functional residuals*", the Joint Statistical Meetings (JSM), Washington D.C. 08/2022
- **Refereed session with travel award**, "*Model diagnostics of discrete data regression: a unifying framework using functional residuals*", Symposium on Data Science and Statistics (SDSS), Pittsburgh, PA. 06/2022
- **Student award presentation**, "*Model diagnostics of discrete data regression: a unifying framework using functional residuals*", New England Statistics Symposium (NESS), Mansfield, CT. 05/2022
- **Contributed session**, "*Analyzing conflicting information via multi-dimensional textual network analysis framework*", INFORMS Annual Meeting, Virtual. 10/2020

TEACHING EXPERIENCE

Teaching Interests

- Introduction level courses: Introduction to Business Analytics, Descriptive Analytics, Predictive Analytics Prescriptive Analytics, Business Intelligence.
- Advanced level courses: Applied Linear Regression, Statistical Modeling, Data Mining, Optimization Methods in Analytics, Big Data Analytics, Probability models.
- Programming courses: AI and Machine Learning Algorithm, Data Wrangling, Text Mining, Statistical Computing, Data Manipulation and Visualization with Python/R.

Independent instructor

— My courses have received high ratings, ***surpassing the departmental evaluation scores****, which typically range from* **6.56** *to* **7.31** *(Note: Evaluation score based on the grouped median of the overall excellence of the course on scale of 1(lowest) - 8(highest).)*

- Online courses:
  - **BANA 4085 Spreadsheet Analytics** (Undergraduate level, Eval: 8.0/8.0) Spring 2021
  - **BANA 7046 Data Mining I** (Graduate level, Eval: 7.3/8.0) Spring 2022

- Hybrid courses:
  - **BANA 6043 Statistical Computing** (Graduate level, Eval: 7.8/8.0) Fall 2021
- In-person courses:
  - **BANA 7025 Data Wrangling** (Graduate level, Eval: 7.6/8.0) Fall 2022
  - **BANA 7046 Data Mining I** (Graduate level, Eval: 7.8/8.0) Spring 2023

Teaching assistant

— *Undergraduate level courses*

- BANA 2081 Business Analytics I
- BANA 2082 Business Analytics II
- BANA 4085 Spreadsheet Analytics
- BANA 4137 Descriptive Analytics and Data Visualization
- BANA 4143 Data Management for Analytics

— *Graduate level courses*

- BANA 6043 Statistical Computing
- BANA 7052 Applied Linear Regression
- BANA 7046 Data Mining I
- BANA 7047 Data Mining II

Second Reader for Capstone Essays

- The Capstone essay forms an essential part of the Master of Science in Business Analytics curriculum, necessitating a comprehensive analysis grounded in a real-world data project.
- I served as the second reader for **14 individual projects**, covering a wide spectrum of topics. These spanned from **forecasting demand and segmenting customers to analyzing customer reviews, classifying toxic comments, and predicting medical claims**.

PROFESSIONAL MEMBERSHIPS

- Member, American Statistical Association (ASA).
- Member, The Institute for Operations Research and the Management Sciences (INFORMS).
- Member, Institute for Mathematical Statistics (IMS).
- Member, New England Statistical Society (NESS).

SKILLS

Programming	R, Python, SPSS, SAS, C++, Matlab, Stata, and Eviews.
R Packages Developed	SurrogateRsq.
Languages	English (fluent); Chinese (native); Japanese (N3).