Eric Zhang

Email: ezhang1218@gmail.com | Personal Website: https://ezhang1218.github.io | Phone: 9148743315

Education

University of North Carolina at Chapel Hill — Chapel Hill, NC
PhD Candidate, Department of Biostatistics — *Aug. 2021 – December 2025 (Expected)*
Research Areas: High-dimensional data, Dimension Reduction, Statistical Modeling, Machine Learning, Data Science.

Cornell University, BA — Ithaca, NY
Double Major: Mathematics, Statistics. Minor: Computer Science. GPA: 3.70 — *Aug. 2017 – May 2021*

Publications

Zhang, E., Li, D. (2025). Contrastive Functional Principal Components Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence, 39(21).* [Link] [Code]

Zhang, E., Love, M., Li, D. (2025). Contrastive CUR: Interpretable Joint Feature and Sample Selection for Case-Control Studies. *arXiv:2508.11557.* [Link]

Muhlebach, M., **Zhang, E.**, et al. (2025). Association Between Inhaled Antibiotic Use and Treatment-Emergent Organisms among Canadian People with CF. *Journal of Cystic Fibrosis.*

Muhlebach, M., **Zhang, E.**, et al. (2024). Changes in factors associated with inhaled antibiotic prescriptions for people with cystic fibrosis over time in the U.S. *Journal of Cystic Fibrosis, 24(1), 98–104.*

Experience

JPMorgan Chase — Jun. 2025 – Aug. 2025
Quantitative Analytics Summer Associate — *New York*

- Applied XGBoost and evaluated survival model approaches on various vehicle lease exits outcome. XGBoost significantly increased AUC, and the survival model achieved a comparable Harrell's C-index, indicating model was successfully ranking risk over time.
- Examined time-varying features in relation to time to default outcomes using Cox proportional hazards models with counting process and generated time-dependent AUCs.
- Evaluated linear mixed effects models on car depreciation data and proposed nested random effects.

UNC at Chapel Hill — Aug. 2021 – Present
Graduate Research Assistant, PhD Candidate — *North Carolina*

Dimension Reduction for Time-Series Data

- Developed a contrastive dimension reduction method to uncover variance unique to or enriched in a group of interest.
- Applied to case-control time-series datasets such as stock prices and gait cycles, achieving improved clustering and higher Silhouette Scores. Revealed insights in complex time-series data that were not captured by baseline methods.

Unsupervised Feature and Sample Selection

- Addressed interpretability issues in PCA by developing unsupervised feature selection and sample selection techniques in a case-control framework.
- Proposed methods allow extraction of interpretable features and samples that drive contrast between groups.

Causal Inference

- Applied inverse probability weighted generalized estimating equations (IPW-GEE) to longitudinal clinical data.
- Estimated causal effect of chronic antibiotic use on outcomes such as lung function and infection presence.

LLM Embeddings with ECG Data

- Artifact detection in ECG data is a complex problem. Signal-to-noise ratio is small, and even physicians have difficulty identifying them. We look to improve detection by fine-tuning multimodal large language models (LLMs) such as Qwen and LLaVA on ECG image data to detect artifacts.

- Extracted Contrastive Language Image Pretraining (CLIP) embeddings and trained downstream classifiers, achieving an AUC of 0.95.

Ernst & Young Jul. 2020 – Aug. 2020

Quantitative Advisory Intern *New York*

- Performed fair value (MTM) valuation for various financial products such as CDS, FX forwards, option, etc.

Awards

National Institude of Environmental Health Sciences

T32 Training Grant in Environmental Biostatistics.

National Chess Master

United States Chess Federation Certified.

Skills

Languages : Python, R, SQL, Pytorch, SAS.

Machine Learning: Andrew Ng's Stanford courses: Machine Learning, Neural Networks & Deep Learning, Hyperparameter tuning, Regularization and Optimization.

Data Science: Data Science: Statistical modeling, Hypothesis testing, Python (e.g.scikit-learn, numpy, pandas, matplotlib)