# NIANLIN CHEN

3369 Primrose Ln, Ypsilanti, Michigan 48197, U.S.

Email: chennianlin26@163.com | Tel: +1 (530)-220-4570

## EDUCATION

**University of Michigan – Ann Arbor**                                                       Ann Arbor, MI

*MS in Applied Statistics*                                                                  Aug 2021 – May 2023

- <u>Main Courses:</u> Linear Regression, Statistical Inference, Survival Time Analysis, Bayesian Modeling, Statistical Computing, Machine Learning, Risk Analysis, etc.

**University of California – Davis**                                                              Davis, CA

*BS in Mathematics* (*Mathematical Analytics & Operations Research*)        Sep 2017 – Jun 2021

- <u>Main Courses:</u> Calculus, Vector Analysis, Linear Algebra, Mordan Algebra, Differential Equation, Real Analysis, Numerical Analysis, Probability Theory, Stochastic Process, Optimization, Time Series Analysis, Math for Data Analytics, Regression Analysis, Analysis of Variance, etc.

## EMPLOYMENT AND INTERNSHIP

***Statistical Programmer***          **EDETEK** (Remote, US)                     Jun 2023 – Present

- Transform clinical study protocols and Statistical Analysis Plans (SAPs) into structured data workflows by mapping raw clinical trial data into Study Data Tabulation Model (SDTM) datasets and conducting analyses on Analysis Data Model (ADaM) datasets via SAS.
- Develop and validate tables, listings, and figures (TLFs) in support of regulatory submissions, adhering to industry standards and project-specific guidelines.
- Perform comprehensive data validation checks on raw clinical databases using SQL, R, and SAS to ensure data accuracy, consistency, and compliance with project requirements; and generate actionable checklists for ongoing quality control.
- Contribute to the continuous improvement of internal coding standards (including SAS Macro optimization), documentation practices, and workflow efficiencies, enhancing reproducibility and efficiency across multiple clinical studies.

***Data Analyst Intern***          **Teva Pharmaceutical** (Remote, US)          May 2022 – Sep 2022

- Reviewed and organized clinical study data, including protocols, SAPs, visualization outputs, SDTM/ADaM datasets to ensure regulatory compliance and data integrity.
- Conducted SAS-based quality checks and data validation procedures to identify inconsistencies, clarify variable definitions, and maintain data integrity.
- Validated ad-hoc and post-hoc queries in SAS, maintaining accuracy and consistency in clinical data analyses.

## RESEARCH ASSISTANTSHIP

**Bradburd Lab, University of Michigan, Ann Arbor**                          Oct 2022 – Mar 2023

- Investigated genetic population structure using continuous and discrete linear models, quantifying ancestry proportions and relatedness decay rates across different geographic distances.
- Employed cross-validation to select the optimal model complexity.
- Leveraged Hamiltonian Monte Carlo (HMC – an algorithm/technique to sample from posterior distribution in Bayesian for efficient exploration) for Bayesian parameter estimation to obtain robust

posterior inference on the selected model.

- Contributed to the design and validation of new R packages aimed at improving reproducibility and analytical efficiency in population genetics research.

**RTG (Research Traineeship Group): Statistics in the 21st Century – Objects, Geometry, and Computing** (NSF-Funded Project, Award No. 1148643, UC-Davis)                    Feb – Jun 2020
*Explored how bootstrap resampling improves variance estimation and stability of principal components in high-dimensional datasets.*

- Designed and executed experiments using bootstrap resampling techniques, systematically adjusting PCA parameters (e.g., number of components, covariance matrix) to identify conditions improving variance estimation and component stability.
- Standardized datasets and conducted numerical experiments in MATLAB and R to assess the performance of bootstrap via PCA approaches.
- Extended simulation study across a fine grid of eigenvalue-decay parameters ($\beta \in \{0, 0.25, 0.5, \ldots, 4.0\}$) with dimensions $n = p = 100$ (and later $n = p = 1000$), and bootstrapped replicates M up to 5000.
- Analyzed 95% bootstrap CI coverage over a range of eigenvalue-decay parameters ($\beta$), and presented the findings through MATLAB visualizations.
- Optimized computational workflow by parallelizing resampling jobs on the Galileo cloud cluster, which not only reduced overall run-time by around 60% but also scaled to larger M and higher p without local hardware limits.
- Synthesized findings into a comprehensive research report, prepared a conference poster, and delivered presentations at the department's student conference.

## SELECTED COURSE RESEARCH PROJECTS

**The Development of an R Package for dPCA**                    Sep – Dec 2022
*Course: Statistical Computing | University of Michigan, Ann Abor*

- Architected and implemented a dPCA R package in a team of three, enabling demixed principal component analysis directly in R for the first time.
- Developed robust data-handling routines to flatten and marginalize multi-dimensional neural population arrays ($D \times T \times S \times N$), leveraging ANOVA-style averaging to disentangle time, stimulus, decision, and mixed components.
- Engineered three reduced-rank regression solvers—Randomized SVD, full SVD, and QR decomposition—allowing users to trade off speed and accuracy; benchmarked on matrices up to 240,000 × 800 to demonstrate up to 60% speed gains with RSVD.
- Incorporated ridge regularization ($\lambda$) into the loss function and exposed it as a user-configurable parameter, improving component stability and reducing over-fitting on simulated datasets.
- Built a simulation suite and plotting functions (using ggplot2) to generate and visualize demixed components across task parameters, producing publication-quality figures for time- and stimulus-driven effects.
- Ensured package quality and ease of adoption by writing comprehensive README, vignettes, and unit tests (testthat), and setting up GitHub Actions for continuous integration and code coverage.
- Deployed the R package as open-source on GitHub, and provided end-to-end examples illustrating installation via devtools, data preparation, and interpretation of dPCA outputs.
- Addressed and presented limitations of standard PCA by improving efficiency and interpretability over

standard PCA.

**Predicting the Risk of Pre-Eclampsia in First-Time Pregnant Women (Nulliparas)**

*Course: Practice And Communication in Applied Statistics | University of Michigan, Ann Arbor*

Sep – Dec 2022

- Investigated risk factors for pre-eclampsia in nulliparous women, emphasizing the clinical importance of early prediction for maternal health.
- Led the Logistic Regression and Random Forest model development, including hyperparameter tuning and evaluation via cross-validation.
- Identified key physiological and psychological factors, including blood pressure, pre-pregnancy weight, stress levels, etc.
- Wrote the parts of methodology and results in the final report, while ensuring consistency and coherence across the whole project.

**Music Genre Classification Using Audio Features**                Jan – Apr 2022

*Course: Statistical Learning II: Multivariate Analysis | University of Michigan, Ann Arbor*

- Preprocessed the Kaggle dataset containing audio features, such as loudness, danceability, tempo, etc. by handling missing values and standardizing numerical features for modeling.
- Investigated feature importance using PCA and correlation heat maps to guide model selection.
- Implemented multiple algorithms, including KNN, Random Forest, AdaBoost, etc., and fine-tuned hyperparameters via grid search and cross-validation.
- Analyzed confusion matrices to uncover class-specific performance, and explained the key features driving classification decisions.
- Summarized findings in a comprehensive project report, including motivations, methods, results, and future directions.

**Bayesian Sales Forecasting Project – Walmart Weekly Sales**                Jan – Apr 2022

*Course: Bayesian Inference | University of Michigan, Ann Arbor*

- Applied Bayesian linear regression and hierarchical Bayesian models to forecast weekly sales for Walmart stores, using historical sales, store-level attributes, and economic indicators.
- Consolidated multiple Walmart datasets, handled missing values, and standardized numerical variables.
- Mapped out correlations, identified skewness in weekly sales, and visualized key trends (e.g., store size vs. sales).
- Developed Bayesian linear regression using PyMC3, specifying prior distributions and running multiple MCMC chains; diagnosed convergence via trace plots and autocorrelation checks.
- Analyzed posterior distributions, and highlighted significant predictors, including unemployment, fuel price, and Customer Price Index (CPI).
- Authored major sections of the final project report, including introduction, data, hierarchical model setup, etc.

## TEACHING ASSISTANTSHIP

***Listed Tutor*, LSA (College of Literature, Science and the Arts), University of Michigan, Ann Arbor**

Aug 2021 – May 2023

- Provided assistance for the students who needed help with statistical concepts, ranging from fundamentals to advanced topics.
- Adapted my teaching approaches to accommodate students with diverse backgrounds.

## VOLUNTEER EXPERIENCE

***Public Health Ambassador*, UC-Davis**                                              Oct 2020 – Jun 2021

- Helped communicate interim public health policies to keep students informed about evolving COVID-19 guidelines.
- Led outreach initiatives to educate the student body on health risks and preventive strategies, fostering a culture of shared responsibility and vigilance.
- Collaborated with campus stakeholders to implement, monitor, and adapt safety protocols, supporting a safe return to in-person learning and activities.

## TECHNICAL SKILLS & CERTIFICATIONS

**Programming Languages:** SAS, SQL, MATLAB, R, Python

**Certifications:**

- *SAS Certified Specialist:* Base Programming Using SAS 9.4
- *SAS Certified Professional:* Advanced Programming Using SAS 9.4

**Data Analytics & Modeling Tools:**

- Python: *Pandas, NumPy, statsmodels, scikit-learn, SciPy*
- R: *caret, xgboost, rstan, rstanarm*

**Data Visualization & Reporting:**

- Python: *Matplotlib, Seaborn*
- R: *ggplot2, lattice*
- Others: *Tableau, Excel, LaTeX, R Markdown*