Guanyi Yang, BSc, BEcon

Master's Student in Applied Statistics

University of Michigan, Ann Arbor, MI, USA

guanyiy@umich.edu

Education

University of Michigan Ann Arbor, MI
Master of Science in Applied Statistics Aug. 2024 - Present
Research Advisor: Emily M. Briceño, PhD

- **Relevant Coursework:** Probability & Distribution Theory, Statistical Inference, Regression Analysis, Advanced Regression, Bayesian Modeling, Statistical Machine Learning, Causal Inference, Data Science in Python
- **Selected PhD-level courses:** Computational Optimization Methods for Statistics
- **Additional Training:** Latent Variable Models & Longitudinal Data Analysis (3-day workshop, Brown University, 2025; instructor: Richard N. Jones, ScD)
- **Current Research:** Multinomial logistic regression for MCI classification (HRS-HCAP, Mex-Cog); Nepal HCAP Cognitive Assessment Adaptation.

University of Liverpool & Xi'an Jiaotong-Liverpool University Liverpool, UK & Suzhou, China
BSc (Honours, First Class) & BEcon in Financial Mathematics Sept. 2020 - Jul. 2024
Advisor: Yichen Liu, PhD; Thomas Selig, PhD

- **Relevant Coursework (statistics & probability focus)**: Linear Algebra, Multivariable Calculus, Real Analysis for Financial Mathematics, Probability & Statistics, Statistical Distribution Theory, Stochastic Processes & Calculus, Linear Statistical Models, Applied Multivariate Statistics, Econometrics I & II
- **Undergraduate Dissertation:** *Mathematical Methods in Statistics and Finance* — explored variational methods for statistical estimation, option pricing, and discrete-probability models for financial applications.
- **Awards:** University Academic Achievement Award (Top 10%) (2022 & 2023), Across-China International Youth Leadership Lecturer Group (Top 1%) (2022), Excellent Student Award (Top 5%) (2022)

Research Experience

University of Michigan, Dept of Physical Medicine & Rehabilitation, Ann Arbor, MI May 2025 -- Present

Research Project: Cross-national Cognitive Aging and MCI Classification (HRS-HCAP, Mex-Cog, LASI-DAD) | R
Role: Research Assistant
Advisor: Emily M. Briceño, PhD;
Collaborators: Miguel Arce Rentería, PhD (Columbia University); Zachary J. Kunicki, PhD (Brown University)

- Managed and analyzed large-scale cognitive and demographic datasets, including the Health and Retirement Study Harmonized Cognitive Assessment Protocol (HRS-HCAP, n ≈ 3,500) and Mexican Cognitive Aging Ancillary Study (Mex-Cog, n ≈ 2,000).
- Harmonized key variables (e.g., rurality, age) across studies by consolidating multiple coding schemes and aligning definitions with prior literature, then derived analytic variables (stroke, hypertension, diabetes and depression) for regression and classification analyses.
- Fit and compared multinomial logistic regression and ordinal regression (PO, PPO, adjacent-category) for the same 3-class cognitive-status outcome (Normal; MCI w/ memory impairment; MCI w/o memory impairment).
- Addressed data challenges, including skewed covariates, class imbalance, and missing data, through transformations, weighting, and sensitivity analyses.
- Produced a 20+ page reproducible R Markdown pipeline and delivered results in weekly lab presentations and formal reports to Columbia and Brown collaborators.

University of Michigan, Institute for Social Research, Ann Arbor, MI May 2025 -- Present

Project: Nepal HCAP Cognitive Assessment Adaptation | R, Mplus
Role: Research Assistant
Advisor: Emily M. Briceño, PhD
Collaborators: Dirgha Ghimire, PhD (UM); Carlos Mendes de Leon, PhD (Georgetown University)

- Performed data management, cleaning, and quality checks on Nepal HCAP pilot data. Engaged in frequent feedback with the field office and data management team in Nepal to improve quality of data and documentation. Handled and evaluated patterns of missingness, skewness and data entry errors. Created protocols for recoding data and generating derived variables.
- Led the analysis of newly adapted cognitive tests in the HCAP protocol by comparing them with standard cognitive test items across education groups, demonstrating reduced floor effects and improved validity.
- Led comparative analyses of adapted cognitive assessment tools for low-literacy populations (matchstick-based visuospatial construction and recall, adapted symbol cancellation test, novel cognitive screening test), applying nonparametric ART ANOVA and two-way ANOVA to evaluate feasibility, education effects, and convergent and discriminant validity.
- Conducted confirmatory factor analysis (CFA) to examine whether the adapted HCAP battery reliably captured the memory domain in older adult populations.
- Co-authored two abstracts submitted to INS 2026, including one as first author.

University of Michigan, Ann Arbor, MI Feb. 2025 - Jul. 2025

Project: Digital Cages Project — High-Dimensional Behavioral State Analysis | R
Role: Research Assistant
Advisors: Ivo D. Dinov, PhD & Simeone Marino, PhD

- Analyzed high-resolution actigraphy data from 48 mice in digitally monitored cages, comprising millions of time-series observations.
- Applied PCA, sparse PCA (NExOS algorithm), k-means, and hierarchical clustering to extract latent behavioral states and identify interpretable daily and circadian patterns.
- Built reproducible pipelines in R for dimensionality reduction, clustering, and visualization of high-dimensional data.
- Addressed data challenges including temporal autocorrelation, high dimensionality, and noise, developing feature selection strategies to improve interpretability.
- Presented preliminary analyses in weekly lab meetings and internal poster sessions, contributing to ongoing methodological development within the research group.

Xi'an Jiaotong-Liverpool University, Suzhou, China Jun. 2024 -- Aug. 2024

Project: Algorithmic and Graph-Theoretic Analysis of Parking Functions | Python
Role: Research Assistant
Advisor: Thomas Selig, PhD

- Applied graph-theoretic methods (bipartite matchings, adjacency matrices, friendship graphs) to characterize parking function structures and derive recurrence relations.
- Built a Python-based simulation framework using NetworkX and NumPy, enabling large-scale validation of combinatorial conjectures.
- Designed novel algorithms for multi-agent parking scenarios, improving computational efficiency by 40% compared to standard enumeration.
- Conducted 100,000+ simulations to analyze structural patterns, validating theoretical predictions through visualization in Matplotlib.
- Results published in Enumerative Combinatorics and Applications, 4(3), Article S2R21 (2024).

Publications

Selig, T., Kang, Y., **Yang, G.**, Zhang, Y., & Zhu, H. (2024). On friendship and cyclic parking functions. *Enumerative Combinatorics and Applications*, 4(3), S2R21.

Manuscripts in Preparation

Yang, G., Ghimire, D., Dhakal, U., Bogati, U., Sapkota, H., Shrestha, L., Pandit, P., Jalan, P., Mendes de Leon, C., & Briceño, EM. *Tentative title: Adaptation of a Visuospatial Construction and Visual Memory Test for an Educationally Diverse Older Adult Population in Nepal.*

Briceño, EM, **Yang, G.**, Bogati, U., Jalan, P., Shrestha, L., Dhakal, U., Sapkota, H., Rai, J., Pandit, P., Mendes de Leon, C., & Ghimire, D. *Tentative title: Development of a Cognitive Screening Test for Educationally Diverse Older Adults in Nepal.*

Jeon, S., Kunicki, Z., Kamalyan, L., Prieto, S., Yang, G., Snider, M., Marshall, L. W., Saenz, J., Mejia Arango, S., Wong, R., Arce Rentería, M., & Briceño, EM. Tentative title: Longitudinal Measurement Invariance of the Harmonized Cognitive Assessment Protocol in the Mexican Cognitive Aging Ancillary Study (Mex-Cog).

Briceño, EM., Adhikari, N., Sapkota, H., Dhakal, U., Jalan, P., Shrestha, L., Sapkota, H., Rai, J., Bogati, U., **Yang, G.**, Prandit, P., Mendes de Leon, C. & Ghimire, D. *Tentative title: Adaptation of the Harmonized Cognitive Assessment Protocol in Nepal.*

Snider, M., Kunicki, Z., Jeon, S., Kamalyan L., Marshall L., Prieto S., Rich, A., **Yang, G.**, Briceño, EM. & Arce Rentería, M. *Tentative title: Cross-national comparison of hospitalization on cognition.*

Conference & Presentations

Yang, G., Ghimire, D., Dhakal, U., Bogati, U., Sapkota, H., Shrestha, L., Pandit, P., Jalan, P., Mendes de Leon, C., & Briceño, E. M. (Accepted for presentation, February 2026). *Adaptation of a visuospatial construction and visual memory test for an educationally diverse older adult population in Nepal.* Poster to be presented at the **International Neuropsychological Society (INS) 54th Annual North American Meeting**, Philadelphia, PA. (First author)

Briceño, EM., **Yang, G.**, Bogati, U., Jalan, P., Shrestha, L., Dhakal, U., Sapkota, H., Rai, J., Pandit, P., Mendes de Leon, C., & Ghimire, D. (Accepted for presentation, February 2026). *Development of a Cognitive Screening Test for Educationally Diverse Older Adults in Nepal*. Poster to be presented at the **International Neuropsychological Society (INS) 54th Annual North American Meeting**, Philadelphia, PA. (Co-author)

Project Experience

University of Michigan, Ann Arbor, MI Sept 2025 – Present

Project: Causal Effect of Oral Health on Hearing (NHANES 2021–2023) | R (tidyverse, WeightIt/survey, tmle3)
Role: Team Research Project (ongoing)

- Estimated the average treatment effect (risk difference) using doubly robust estimators with overlap diagnostics (propensity score distributions; weight truncation at 1 and 99 percent).
- Achieved covariate balance (maximum standardized mean difference improved from 0.23 to 0.06); conducted sensitivity under propensity score restriction 0.05 to 0.95.
- Preliminary result: risk difference +4.5 percentage points (95% CI +1.8 to +7.2); trimmed analysis +3.6 percentage points; documented partial-overlap regimes where estimates should not be trusted.
- Delivered one-click reproducibility (fixed seeds and environment) and a negative results note; presented design and interim results in class with faculty Q&A.

University of Michigan, Ann Arbor, MI Sept 2025 – Present

Project: Modeling Time Effects on Bike-Share Counts (Washington, DC, 2011–2012) | R (MASS, mgcv, lme4, caret)
Role: Team Research Project (ongoing)

- Implemented a model ladder: Poisson, Quasi-Poisson, Negative Binomial, generalized additive model with a time spline, and a mixed model with month and weekday random effects.
- Used cross-validated RMSE, out-of-sample log-likelihood, Pearson and deviance residuals, and leverage and influence diagnostics; compared zero-inflated models when the zero rate exceeded 12 percent.
- Reduced cross-validated RMSE from approximately 800 to approximately 620 (about 20 to 25 percent improvement); out-of-sample log-likelihood improved by 14 percent versus Poisson.
- Presented findings in class and addressed questions on dispersion and overfitting; submitted a final report with scripted end-to-end reproduction.

University of Michigan, Ann Arbor, MI Sept. 2024 -- Dec. 2024

Project: Bayesian Hierarchical Analysis of Work Hour Trends Across Sectors | R (rstan)
Role: Individual Course Project
Instructor: Yixin Wang, PhD

- Developed a Bayesian hierarchical model for work hour data across 20+ industries, refining prior specification and adaptive sampling strategies to improve parameter estimation accuracy.
- Implemented MCMC sampling (5,000 iterations) with step-size tuning and parallelized computation, achieving stable convergence ($\widehat{R} < 1.01$) and reducing runtime by 30%.
- Validated model consistency by comparing closed-form posterior approximations with simulation-based estimates; improved predictive accuracy by 15% using LOO-CV and DIC.
- Generated 50+ industry-specific visualizations in ggplot2, highlighting sectoral differences (e.g., high variability in retail vs. stable patterns in IT and education).
- Produced a technical report and presented findings in class, linking statistical insights to post-pandemic workforce policy discussions.

Xi'an Jiaotong-Liverpool University, Suzhou, China Sept. 2023 -- Jul. 2024

Project: MVP Parking Functions in Abelian Sandpile Models | Java
Role: Undergraduate Researcher
Instructor: Thomas Selig, PhD

- Conducted a literature review on Maximum-Value Priority (MVP) parking functions, focusing on their combinatorial properties and applications in self-organized criticality and resource allocation.
- Extended the Abelian Sandpile Model (ASM) to incorporate MVP parking function constraints, formulating new transition rules and recurrence relations.
- Implemented Java-based simulations to analyze the stability and dynamics of ASM configurations under MVP parking function conditions.
- Evaluated simulation outcomes with statistical and algebraic methods, assessing the feasibility of integrating ASM with MVP parking functions in discrete dynamical systems.

Xi'an Jiaotong-Liverpool University, Suzhou, China Sept. 2023 -- Jan. 2024

Project: Numerical Simulation and Hedging Strategies for European Options | Python
Role: Individual Course Project
Instructor: Yi Hong, PhD

- Implemented Euler–Maruyama discretization to approximate stochastic differential equations for European option pricing, simulating 100,000+ asset price paths under geometric Brownian motion.
- Designed and optimized Delta-neutral and Delta–Gamma-neutral hedging strategies, integrating Black–Scholes closed-form solutions and assessing sensitivity to volatility regimes.
- Evaluated simulation accuracy and numerical stability, reducing pricing error to 1.8% deviation from Black–Scholes benchmarks across varying time steps and path counts.
- Quantified hedging performance through replication error analysis, visualizing risk exposure and convergence trends using Python (NumPy, Matplotlib).

Skills

- **Programming:** R (tidyverse, rstan, ggplot2), Python (pandas, scikit-learn, PyTorch), MATLAB, Java, Mplus
- **Statistical & ML Methods:** Regression (linear, logistic, multinomial), Bayesian inference (MCMC, hierarchical models), Regularization (Lasso, Ridge), Tree-based models (Random Forests, Gradient Boosting), SVM, Multivariate analysis (PCA, clustering, factor analysis), Dimension reduction (t-SNE, UMAP), Causal inference, Nonparametric tests
- **Tools & Communication:** SQL, Git/GitHub, LaTeX, R Markdown/Quarto; data visualization & academic reporting