

Boyuan Zhang

E-mail: boyuan.zhang@nyu.edu | Website: gary-boyuan-zhang.github.io

Education

New York University, College of Arts and Sciences New York, NY
Bachelor of Arts, Computer and Data Science Jan. 2021 – Dec. 2023(exp.)

- Minor in Mathematics and Business Studies
- Major GPA: 3.9 / 4; Cumulative GPA: 3.82 / 4
- *Honor: Dean's List (Academic years 2021-23)*

The Experimental High School Attached to Beijing Normal University (ESBNU) Beijing, China
Sep. 2017 – June 2020

- *Honor: Golden-Ship Award (Top 1%)*

Syracuse University, Falk College of Sport and Human Dynamics Syracuse, NY
Berlin Sports Analytics Academy July 2019

Academic Interest

Causal discovery and inference; Statistical machine learning; Bayesian and Statistical inference;
Applications in healthcare, public policy, social science, and sports.

Publications and Preprints

1. **B. Zhang** and Wesley Tansey. "Testing for Individual Causal Effects in Observational Data." *In Preparation.*
2. W. Herlands, **B. Zhang**, E. McFowland III, and D. B. Neill. "Automated Discovery of Difference-in-Differences." *In Preparation for submission to Journal of the American Statistical Association.*
3. **B. Zhang**, E. T. Hoac, and P. Hoang. "A Regularized Adjusted Expected Goal Plus-Minus Model for Soccer." *In Preparation.*

Research Experience

Memorial Sloan Kettering Cancer Center, Dept. of Epidemiology and Biostatistics New York, NY
Machine Learning for Cancer Lab June 2023 – Present

- Research Area: Causal Testing, Empirical Bayes, Dose-response Modeling, Drug Discovery
- Supervisor: Dr. Wesley Tansey
- Developed a nonparametric compliance testing procedure for individual causal effects from observational data with potential confounding and non-compliance under an additive noise model
- Demonstrated effective control of the false discovery rate at the target level in simulations while maintaining reasonable power and providing more accurate estimates of the average treatment effect compared to BART, Causal Forest, and FDR Regression
- Discovered biologically rational markers that predispose a cell line to be sensitive or resistant to treatment with a targeted agent using data from a large cancer drug study

New York University, Courant Institute of Mathematical Science New York, NY
Machine Learning for Good Lab Feb. 2023 – Present

- Research Area: Causal Discovery and Inference, Subset Scanning, Machine Learning
- Supervisor: Dr. Daniel Neill

- Refined and improved the subset scanning algorithms for difference-in-difference discovery
- Identified both heterogeneous treatment discontinuities in arbitrary-dimensional categorical data and appropriate control subsets with statistical and econometric validity
- Demonstrated robust performance in simulations, and discovered novel policy-relevant insights in complex public policy datasets

Carnegie Mellon University, Department of Statistics and Data Science Pittsburgh, PA
 Summer Undergraduate Research Experience in Statistics June 2022 – July 2022

- Research Area: Data Science, Sports Analytics, Player Evaluation, Plus-Minus Models
- Supervisors: Dr. Ronald Yurko, Dr. Konstantinos Pelechrinis
- Developed a new regularized adjusted plus-minus model for evaluating player's performance and contribution in soccer
- Incorporated priors learned from box score statistics into a regularized regression framework
- Demonstrated better predictability and interpretability than existing methods through simulations

Peking University, School of Electronics Engineering and Computer Science Beijing, China
 Science Talent Program Mar. 2018 – Dec. 2018

- Research Areas: IoT, Embedded Systems, Computer Electronics, and Information Technology
- Supervisor: Prof. Yanjun Yang
- Designed a personalized intelligence mailbox based on IoT technology
- Developed mobile app in Android Studio to control the mailbox remotely via Gizwits Cloud
- Assembled the cabinet and utilized Arduino to manage modules and functions on the mailbox

Teaching Experience

New York University, Center for Data Science New York, NY
 Teaching Assistant May 2023 – Present

- *DS-UA 301: Advanced Techniques in ML and Deep Learning*. Prof. Parijat Dube. Fall 2023
- *DS-UA 201: Causal Inference*. Prof. Parijat Dube. Summer 2023

New York University, Courant Institute of Mathematical Science New York, NY
 Teaching Assistant Sep. 2022 – Present

- *CSCI-UA 2-009: Intro to Computer Programming*. Prof. Matthew Zeidenberg. Fall 2023
- *CSCI-UA 2-004: Intro to Computer Programming*. Prof. Mihir Patil. Spring 2023
- *CSCI-UA 2-006: Intro to Computer Programming*. Prof. David Prager. Fall 2022
- *CSCI-UA 2-010: Intro to Computer Programming*. Prof. Khye Borg Liew. Fall 2022

The Experimental High School Attached to Beijing Normal University Beijing, China
 Instructor Sep. 2019 – Dec. 2020

- *General Elective: Orienteering*

Talks

QSURE Symposium New York, NY
Testing for Individual Causal Effects in Observational Data Aug. 2023

Carnegie Mellon Sports Analytics Conference Pittsburgh, PA
A Regularized Adjusted Plus-Minus Model in Soccer with Box Score Prior Oct. 2022

Carnegie Mellon Sports Analytics Camp	Pittsburgh, PA
<i>A RAPM Model for Soccer Player Ratings (with Edvin Tran Hoac and Phong Hoang)</i>	Aug. 2022
Science Talent Program Computer Science Forum	Sichuan, China
<i>Personalized Intelligence Mailbox Based on IoT</i>	Nov. 2018

Selected Course Projects

Analyzing the Impact of Subway Major Incidents on Passenger Flows	Summer 2023
Comparative Analysis of Collaborative Filtering Algorithms for Recommendation Systems	Spring 2023
Calibrated NBA Win Probability Estimation	Spring 2023
Fingertips Position Estimation of a Robot Hand	Fall 2022
Lip Reading Word Classification	Fall 2022

Work Experience

Learnjoy Sports	Beijing, China
<i>Events Management and Operation Intern</i>	Sep. 2020 – Dec. 2020

Services

NYU Teaches	New York, NY
<i>Academic Tutor</i>	Oct. 2021 – Present
ESBNU Orienteering Varsity	Beijing, China
<i>Captain</i>	Sep. 2017 – June 2020
ESBNU Student Council	Beijing, China
<i>Director of Sports Department</i>	Nov. 2017 – Nov. 2019

Awards

Honorable Distinction in the Computational and Algorithmic Thinking Competition	2019
Second Prize in Beijing Adolescents Science and Technology Innovation Contest	2019
Honorable Mention in High School Mathematical Contest in Modeling	2018
Second Prize in the National Olympiad of Informatics in Provinces	2018

Skills

Coding: Python (TensorFlow, PyTorch), R, SQL, Java, C/C++, MATLAB, Git, Markdown, LaTeX
Tools: Tableau, MongoDB, Android Studio, Arduino, OCAD, Excel, PowerPoint, Word

Relevant Coursework (* graduate-level)

Probability and Statistics:

Mathematical Statistics*, Applied Statistics*, Probability and Statistics 2*, Intro to Stochastic Processes*, Causal Inference, Probability and Statistics, Regression and Forecasting Model

Statistical and Machine Learning:

Statistical Learning Theory*, Theory of Deep Learning*, Machine Learning for Healthcare*, Advanced Techniques in Machine Learning and Deep Learning, Intro to Machine Learning, Predictive Analytics

Data Science: Data Management and Analysis, Responsible Data Science, Intro to Data Science

Mathematical Foundations: Real Analysis, Linear Algebra, Multivariable Calculus, Discrete Math

Computer Science Foundations: Algorithms, Computer Systems Organization, Data Structures