Zhejia Dong

zhejia_dong@brown.edu | +1 401 263 5252 | LinkedIn | GitHub | 121 South Main St, Providence, RI, 02903, USA

Education

Brown University Providence, USA

Master of Science in Biostatistics *Aug. 2021 – May 2023*

- GPA: 4.0/4.0 (Fall 2021, and Spring 2022)
- Relevant Coursework: Linear Models (A), Generalized Linear Models (A), Statistical Inference I (A), Statistical Inference II (A), Bayesian Statistical Methods (A), Statistical Programming with R (A)

Xi'an Jiaotong University (XJTU, C9 League) Xi'an, China

Bachelor of Science with Honors in Mathematics and Applied Mathematics *Sep. 2018 – July 2021*

- Overall GPA: 88.1/100 | Major GPA: 90.1/100 | Rank: 1/33
- Thesis: Robust Estimation for linear model with Response Missing at Random (Best Undergraduate Thesis Award, 1/121), *Advisor: Prof.* Liya Fu
- Relevant Coursework: Functions of Real Variable (100), Advanced Mathematics (98), Advanced Algebra and Analytic Geometry (96), Probability (96), Functional Analysis (96), Mathematical Modeling (96), Data Analysis and Statistical Software (95), Mathematical Statistics (94), Abstract Algebra (93), Differential Geometry (92)
- Honors: Outstanding Graduates, The First Prize Scholarship (2018, 2019, 2020)

Major in Public Policy and Administration *Aug. 2017 – Aug. 2018*

- Overall GPA: 88.7/100 | Major GPA: 90.6/100 | Rank: 1/38
- Relevant Coursework: Microeconomics (90), Macroeconomics (88)

Research Experience

Spurious association due to network dependence with multiple clusters Providence, USA

Avenir Graduate Research Assistant, *Jun. 2022 – Present*

Prof. Youjin Lee *and Prof.* Ashley Buchanan (University of Rhode Island)

- Empirically examined the problem of spurious associations problem under different network structures and two autocorrelation processes
- Investigated the impact of multiple independent clusters on the problem of spurious associations due to network dependence
- Found that spurious associations are more pronounced when each variable is autocorrelated to form a small number of communities within clusters due to network dependence
- Identified that as the number of independent clusters increases, the problem due to network dependence can be reduced

Bias amplification of instrumental variable with unmeasured confounding Providence, USA

Master's Thesis, Prof. Youjin Lee *Nov. 2021 – Present*

- Concentrated on the bias amplification problem due to conditioning on an instrumental variable in a linear model when there are unmeasured confounders between the exposure and outcome variables
- Identified the amplified bias in a new data generating process, in which an instrumental variable is valid conditional on measured confounders
- Derived the amplified bias formula without a strong linearity assumption of the conditional expectation of the unmeasured confounders; proved the estimator from linear regression model that adjusts for a valid instrumental variable would result in greater bias than the estimator that does not adjust for a valid instrumental variable
- Performed numerical experiments to verify the amplified bias from the derived formula

Robust estimation for linear model with response missing at random Xi'an, China

Bachelor's Thesis, Prof. Liye Fu *Oct. 2020 – Aug. 2021*

- Proposed new double robust estimators for linear regression coefficients and population mean when responses are missing at random, through augment inverse probability weighting; proved the proposed estimators are asymptotically normal
- Derived the two estimators from Robit model as the as the propensity score model; applied the EM algorithm to derive the estimated propensity scores; applied the generalized methods of moments to derive the estimated covariate balancing propensity scores

- Performed numerical experiments to compare the performance of proposed estimation methods with the method, that augment inverse probability weighting with logistic regression to estimate both the propensity score and covariate balancing propensity score; the simulation shows our proposed estimator is more robust when both the outcome and propensity score methods are misspecified

Analysis of M-estimation objective functions Xi'an, China

Research Assistant *Nov. 2019– June 2020*

- Studied the multiple M-estimation methods when errors follow normal, t, Cauchy, mixture distributions, and high leverage cases; compared the performance of different objective functions,such as Huber, Hampel, and Welsch, to estimate the parameters of interest.
- Implemented a data-driven procedure that automatically adjusts the value of the tuning constant to provide robustness against outliers; performed numerical experiments to evaluate the performance of different objective functions with or without data-driven procedure through bias and MSE
- Applied the proposed methods to the real datasets, further compared the performance of different methods, and concluded that Welsch and Tukey functions and Huber with data-driven approach have best performance in most cases and are consistent across different sample sizes

Conference/workshop with Presentation

URI/HAIW Collaborative Research Conference. University of Rhode Island *Oct. 2022*

Conference/workshop attendance

2022 Joint Statistical Meetings. Washington DC. *Aug. 2022*

Work Experience

Tencent Technology (Beijing) Co., Ltd Beijing, China

Data Analyst Intern *May – July 2020*

- Obtained the number of thumbs, fans, answers, articles, and other data of Tencent's 55 official accounts on Zhihu by Crawler through *rvast* package in R; conducted word segmentation and word cloud analysis for all articles and answers ($\sim$ 8.6 million words) for data visualization
- Conducted the Spearman Test for correlation of the rare word percent with the popularity of the article and verified the hypothesis that easier-to-understand articles and answers were more popular in Zhihu; summarized results as a completed data analysis report suggesting optimizing proposals for the operations department
- Assisted the Tencent medical team to perform data analysis for COVID-19, obtained data from national and provincial CDC, and drew real-time maps, bar charts, and pie charts to visualize the number of confirmed, suspected, dead, and recovered cases

Extracurricular Activities

Founder and Leader, Living Water Welfare for COVID-19, Wuhan *Feb. – Mar. 2020*

- Launched a public welfare organization and donated medical supplies to the township hospitals in Hubei Province and Wuhan City at the outbreak of COVID-19
- Raised 92,976 RMB in total, bought 9 batches of goggles, alcohol, protective suits, masks, and other medical supplies, and donated all supplies to 23 township hospitals and health centers.

Undergraduate Teaching Assistant, XJTU

- Linear Algebra 2018 Fall, Advanced Mathematics 2019 Fall, Probability and Statistics 2021 Fall

Skills

Programming: R, Python, MATLAB, STATA, SPSS, C++, Java, LaTeX

Standardized Tests: GRE (V164, Q170, AW4.5), TOEFL (107)

Language: English (Proficient), Chinese (Native)