

RENJIE WEI

+1(347)330-0229 ◊ New York, NY

rw2844@cumc.columbia.edu ◊ [Linkedin](#) ◊ [GitHub](#) ◊ [website](#)

EDUCATION

P.h.D in Biostatistics, Columbia University Mailman School of Public Health 2023 - 2027

Master of Biostatistics, Columbia University Mailman School of Public Health 2020 - 2023

Biostatistics Methods I&II, Probability for Biostatistician, Data Science I&II, Topics in Advanced Statistical Methods, Graphical Model, Clinical Trial Methodology, etc.

Bachelor of Medical Laboratory Technology, Major, Peking University 2016 - 2020

Biostatistics, Linear Algebra, Advanced Mathematics, Anatomy, Epidemiology, etc

Bachelor of Economics, Minor, Peking University 2016 - 2020

Probability Theory & Statistics, Intermediate Microeconomics, Intermediate Macroeconomics, Econometric, etc.

RESEARCH EXPERIENCE

Research Assistant July 2022 - Present

Columbia Mailman School of Public Health *New York, NY*

- **Advisor:** Prof. Min Qian, Prof. Bin Cheng
- Develop general study designs of sequential, multiple-assignment, randomized trials (SMARTs)
- Generalize a gate-keeping test for selecting adaptive interventions (AIs) under general SMARTs
- Conduct simulation studies to verify the validity of the gate-keeping test

Research Assistant May 2019 — July 2020

Peking University *Beijing, China*

- **Advisor:** Prof. Bin Cui, Prof. Cunling Yan
- Collect clinical data from electronic medical records about patients from the department of thoracic surgery.
- Clean the data, conduct exploratory data analysis (EDA) and basic statistical analysis.
- Use Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), Random Forest (RF), etc. to build a diagnosis model for the Chinese population. [[Patent Link](#)]
- Contribute to a research review. [[Review Link](#)]

PROJECTS

Bayesian Modeling of Hurricanes Build a Bayesian Model with Markov Chain Monte Carlo algorithm to predict the wind speed for hurricanes and study how the hurricanes related to death and financial loss. [[GitHub Repository](#)]

Predicting Hepatitis C Virus Patient Using Multiple Classification Methods Create a predictive model that performs early detection of Hepatitis C and other liver diseases would allow people to quickly and easily determine their risk/get treatment. [[GitHub Repository](#)]

Simulation Study to Compare Three Survival Models Design a simulation study to compare the accuracy and efficiency of the estimated treatment effects from the three models under various baseline hazard functions and evaluate their robustness against misspecified baseline hazard functions. [[GitHub Repository](#)]

Breast Cancer Diagnosis with Newton-Raphson And Coordinate Descent Lasso Build a predictive model based on logistic regression to facilitate cancer diagnosis, based on Newton-Raphson algorithm and path-wise coordinate-wise LASSO algorithm [[GitHub Repository](#)]

Breast Cancer Diagnosis with Newton-Raphson And Coordinate Descent Lasso Perform EDA to the original data and implement a prediction model of prices. The impact of GDP on local price and consumption of avocado was also assessed. Included interactive visualization tool that will enable people to purposefully check on time-base or region-base information. [[Project Website](#)]

HONORS & AWARDS

- Outstanding Graduates, Peking University 2020
- National Scholarship of Peking University, Minister of Education of China (about \$1200) 2017
- First Prize Scholarship of Outstanding Students, Peking University (about \$1000) 2017 - 2019

PROFESSIONAL EXPERIENCE

Data Analyst June 2023 - September 2023
Center for Behavioral Cardiovascular Health, Columbia University *New York, NY*

- Refine the patient enrollment procedure for the Remote Patient Monitoring for Hypertension (RPM-HTN) study.
- Enroll the patients for a following randomized study and conduct EDA and descriptive analytics on baseline data
- Discover the pattern of patients' home blood pressure monitoring data
- Model the effect of sociodemographic variables on referral status, home blood pressure monitoring pattern, and patient outcome
- Calculate the Treatment Intensity Score (TIS) from prescriptions from electronic health records (EHR) for hypertension medicine

Algorithm & Data Analyst Intern January 2021 - November 2021
Yiducloud *Beijing, China*

- Restructure data from regional data lake using Spark (Ning Bo, Zhe Jiang, China).
- Conduct Real World Research (RWR) based on regional data
- Develop lung cancer risk assessment model using regional data and distributed modeling methods (Ning Bo, Zhe Jiang, China).
- Develop an algorithm calculating personalized life expectancy based on a proportional hazard model derived from 10 million citizens' data and life table (Ning Bo, Zhe Jiang, China). [\[Patent Link\]](#)
- Use SQL to apply statistical analysis and write analytical reports.
- Develop a casual model to evaluate the effect of metformin on the risk of cancer among patients with type 2 diabetes mellitus.

Data Analyst Intern September 2020 - October 2020
PKU Healthcare IT Co., Ltd. *Beijing, China*

- Clean and analyze patient medical records provided by clinicians in neurology.
- Use machine learning method to develop Parkinson's disease prediction model based on data from clinicians and patients' MDS-UPDRS rating scales.

SKILLS

Technical Skills Python, R, Bash, Linux, Spark, MySQL, Git, L^AT_EX, Java, Markdown, PHP
Languages English, Chinese, Japanese