

The Biostatistics PhD program at Columbia University aligns with my academic and career goals: to broaden my statistical toolbox, to refine my research skills, and to deepen my knowledge of biostatistics, epidemiology and public health.

My research journey in Biostatistics began with a project on copula modeling with Professor Evan Ray. Copula is a joint function used to measure the dependence between random variables and widely used in times series modeling. I contributed to the `ncopula` package in R to assist my professor in developing forecasting models. Specifically, I calculated the probability density function and cumulative distribution function to estimate parameters of nested Archimedean copulas with maximum log likelihood. In addition, I included supplementary functions to transform the parameters so that they are within the given bounds for different copula families. This research greatly improved my computation skills as I wrote reproducible code and implemented object-oriented programming with S3 classes to represent nested Archimedean copulas and to perform the estimation.

With this research in statistical methodology under my belt, I continued my journey in the Biostatistics Department at Memorial Sloan Kettering Cancer Center this past summer. Doctor Audrey Mauguen and I investigated the association between biomarker serum bilirubin and survival in Primary Biliary Cirrhosis (PBC) with Cox Proportional Hazards Model, Time-Dependent Cox Model, and Joint Model for longitudinal and time-to-event data. I produced data visualizations with Kaplan Meier curves and spaghetti plots, ran log-rank tests to compare the group survivals, and compared the estimated hazard ratios from these statistical approaches. For the Joint Model, in addition to assessing the proportionality assumption in the survival submodel, I examined the correlation structure in the longitudinal submodel to detect any suspicious multicollinearity. I presented my work at the MSK departmental symposium, the MHC Learning through Applications symposium, and the Electronic Undergraduate Statistics Research Conference.

Motivated by applications of statistics in healthcare, I am expanding the project and incorporating it into my senior thesis under Professor Marie Ozanne's supervision. In addition to documenting my prior work, we plan to revisit the association between serum bilirubin and survival in PBC with the cause-specific hazard model to account for the competing risk. We also hope to explore elastic net regularization as a method for variable selection in survival analysis. In this project, I was fascinated by the pattern mixture models (PMM) and its advantages over the linear mixed-effects models (LME) when doing literature review. PMM reduces potential bias in the estimation by accounting for situations where data are missing not-at-random. Currently, I would like to understand more deeply the possible implementation of PMM in joint modelings and explore if the issue of under-identification in PMM raises any estimation challenges for the joint models.

Columbia University offers an excellent opportunity and a professional training for me to become a scholar in my field of biostatistic. With my strong record, research experience and a strong determination to unravel complicated health problems, I am committed to making a positive contribution to the University's research.