Bin Yang

(757) 707-4088 | by2303@cumc.columbia.edu
New York, NY

EDUCATION

Columbia University; New York, NY   *Expected M.S. June 2022*
- Master of Science in *Biostatistics*, Department of Biostatistics

College of William and Mary; Williamsburg, VA   *Sep 2015 - Dec 2019*
- Bachelor of Science in *Applied Mathematics*, Department of Math

RESEARCH EXPERIENCE

Department of Mathematics, College of William and Mary   **Williamsburg, VA**
Research Assistant   *Aug 2019 - Dec 2019*
Supervisor: Greg Hunt, Ph.D.
- Cleaned and preprocessed the microenvironment microarray data from over thousands of microenvironmental characteristics features; generated features matrices.
- Performed the hierarchical clustering on all feature matrices; utilized the heatmap to assess the data patterns and missingness; imputed the missingness using a modified missing-indicator method.
- Performed principle component analysis to assess the potential clustering pattern for different microenvironment and data quality.s

Raymond A. Mason School of Business, College of William and Mary   **Williamsburg, VA**
Research Assistant   Jan 2019-Aug 2019
Supervisor: Yu Amy Xia, Ph.D.
- Collected the transportation data from the US transportation department and merged high dimensional datasets with complex structure in R to form a new database.
- Investigated the association between the average revenue of logistics companies and emission factors after adjusting for interested covariates based on linear regression.
- Built a final model based on step-wise selection method. Evaluated the functional form of the response.
- Investigated the time effect of association between the annual revenue and emission factors based on GEE model. Selected working correlation structure based on QICs.
- Predicted the subject-specific annual revenue based on given emission factors as well as covariates based on generalized linear mixed model.

PROJECTS

Analysis of Medical Costs Personal Datasets in R
- Performed univariate analysis of the respondents' gender, age, smoking history, region of origins and medical cost.
- Selected the significant predictors for medical cost based on the score statistics of the linear regression model.
- Assessed the functional form of the significant predictors based on f statistics of ANOVA.
- Evaluated the goodness of fitting of the final model based residual analysis.

ACTIVITIES, AWARDS AND HONORS
- Participated in the 2018 EXREEMS-QED summer program at the college of William and Mary.
- Exhibited Poster at the 25th Annual Science Symposium hosted by the college of William and Mary.
- Volunteered at the 2019 Math Awareness Conference at ODU.

SKILLS AND COURSES
- Coding Languages: Python (3 years), R (3 years), Java (1 year), SQL(1 year), Stata (1 year).
- Highlighted courses: Topics in Statistics: Data Mining, Linear Algebra, Statistical Data Analysis, Probability and Statistics of Scientists, Data Structure, Biostatistics Method, Data Science.