Yimeng Shang

☎ +1 (646) 704-5390 ys3298@cumc.columbia.edu https://ys3298.github.io/

EDUCATION

Columbia University, Mailman School of Public Health — New York City, NY, USA
- Master of Science in Biostatistics (GPA: 4.0/ 4.0) — 08/2019-Expected in 05/2021

East China Normal University — Shanghai, China
- Bachelor of Science in Mathematics and Applied Mathematics — 09/2015-06/2019

University of California, Berkeley — Berkeley, CA, USA
- Berkeley International Study Program — 08/2017-05/2018

SKILLS & RELEVANT COURSES

- **Relevant Courses:** Biostatistical Methods, Statistical Inference, Statistical Computing, Data Science, Machine Learning, Epidemiology, Real Analysis, Probability Theory, Functional Analysis, Complex Analysis, Mathematical Modeling
- **Computer Skills**: R, SAS, Python, MATLAB, Microsoft Office Software.

INTERNSHIPS

Eli Lilly China — Shanghai, China, 09/2018– 06/2019

- Data Science& Solution Intern
- Supported data management in clinical trials including data cleaning and missing data query under supervision of China DSS team.
- Participated in the weekly meeting on discussion of statistical methods applied for clinical trials design.
- Constructed quantitative analysis of possible interference risks during clinical trials and organization operation, developed an automatic web page using *ShinyApp* for reproducible monthly analysis to improve efficiency.

RESEARCH/ PROJECTS EXPERIENCE

Implementation and optimization of algorithms on breast cancer diagnosis dataset — 03/2020

- Course: The Advanced Topics in Statistical Computing
- Built a predictive model based on *Logistic Regression* to facilitate cancer diagnosis
- Trained *Logistic Regression* models with *Newton Raphson*, *Gradient Decent* algorithms from scratch in *R* (No package used).
- Decreased the misclassification rate by 4%, via implementing a *Logistic-LASSO Regression* model with *Path-wise Coordinate Descent.*

Gender Pay Gap Claims by Female Doctors at Houston College of Medicine — 12/2019

- Course: Biostatistical Method I
- Used *R programming* to visualize the given data and realize *multiple linear regression* to quantify associations between salaries and gender, as well as any other covariates included in the data.
- Tested confounders and interaction terms between gender and other covariates to build the final model and do related stratification analysis, constructed final report.

The Suicide Rate Project — 11/2019-12-2019

- Course: Data Science I
- Used *R programming* to produce time-based graphics, indicating the relationship between the number of suicide cases and different risk factors, based on the data collected from 13 different sources.
- Implemented *multiple linear regression* with *LASSO* and *BIC criteria* to find the importance of features.
- Made a Crude Suicide Rate map, allowing users to toggle between different years and genders, to demonstrate the changes in the suicide rate over time and visualize regional differences across the U.S.

Establishment and Analysis of Epidemic Model based on the 2014 Ebola Virus Outbreak — 11/2018- 06/2019

- Independent Study, supervised by Professor Ping Bi
- Used *R programming* to visualize the number of infections data across time, indicating number of infections, death toll and infectious victims based on infection cases and death data. And realized function fitting via linear regression.
- Described Ebola outbreak with *SEIR differential equation dynamic system model*, studied existence and stability of balance point with *MATLAB* and verified solution stability in the method of numerical simulation

Statistical Methods in Medical Research and Development — 09/2018- 05/2019

- Independent Study, supervised by Professor Jin Xu
- Attended weekly seminars related to statistical methods of medical journals for clinical trials at one of the biggest local hospital. And reproduced learnt statistical analysis techniques (*e.g. ANOVA, Chi-square test*) with R programming on given clinical data.