

Xiaoyu (Sean) Zhou

202-808-1335 | seanzhou991207@gmail.com | San Diego, CA | [Linkedin](#) | [GitHub](#)

EDUCATION

UC Berkeley, Statistics

Berkeley, CA

Master of Arts

May 2024

- Cumulative GPA: 3.7/4.0
- Relevant Coursework: Statistical Computing, Natural Language Processing, Optimization Models in Engineering, Advanced Probability/Statistics, Longitudinal Data Analysis

William & Mary, Computational Applied Math & Statistics Major, Finance Minor

Williamsburg, VA

Bachelor of Science

December 2022

- Major GPA: 3.94/4.00, Cumulative GPA: 3.82/4.00
- Relevant Coursework: Statistical Learning (Graduate), Bayesian Econometrics, Design of Experiments (Graduate)

SKILLS

- **Programming Languages:** Python (Pytorch, NumPy, SciPy, pandas, matplotlib, seaborn, PyMC3, TensorFlow, Dask), R (caret, ggplot2, tidyverse, gee, lme4), Git/GitHub, Unix/Shell, C/C++, SQL, Stata, Matlab, Minitab
- **Languages:** Fluent in Mandarin, English; Intermediate proficiency in Spanish

PROFESSIONAL EXPERIENCE

USC Alzheimer's Therapeutic Research Institute

San Diego, CA

Biostatistician I

July 2024 - Present

- Perform data validation, diagnostic tests discovery, and prediction modeling related to the A4 trial and the AlzMatch study

W&M University Advancement

Williamsburg, VA

Data Analyst, Stewardship and Donor's Relations Office

September 2021 – December 2022

- Analyzed donor luncheon attendance and donor loyalty using fixed-effects regression in *R/Stata*, revealing an 18% increase in annual donations from luncheon attendees
- Streamlined 12 years of donor data using *pandas/NumPy*; automated the text cleaning process of the 2021/2022 W&M annual donor report using *Python*; presented demographic insights and donation trends to senior endowment directors using visualizations from *matplotlib*
- Helped W&M in becoming the No. 1 public university in the U.S. for alumni giving in 2022

PROJECTS

Weill Institute for Neurosciences at UCSF, Researcher

August 2023 – February 2024

- Analyzed Multiple Sclerosis patients' walking patterns in the SPI2 trial, identifying a crucial 2300 daily step threshold for high disability patients through regression tree analysis
- Employed logistic regression to show increasing daily steps by 800 could reduce disability worsening odds by 22% in highly or moderately disabled patients, integrating Fitbit data, demographics, and questionnaire metrics
- Combined random forests and logistic regression in *R caret* to predict disability worsening with 82% sensitivity, pinpointing baseline EDSS scores and step counts as critical predictors. Applied SMOTE to address the class imbalance bias

gavs: Genetic Algorithm for Variable Selection

August 2023 – December 2023

- Developed a genetic algorithm for variable selection in linear regression and GLMs, tailored for high-dimensional datasets
- Implemented standard genetic operators such as crossover and parent selection while also allowing for user-specified operators and evolution structures; utilized AIC for model evaluation
- Employed *pytest* for rigorous testing of computational components and overall algorithm efficacy

The Charles Center at W&M, Summer Researcher

June 2022 – September 2022

- Utilized synthetic control and lasso regression for a policy analysis on China's two-child policy, constructing a "counterfactual China" from a six-country pool to evaluate policy outcomes
- Applied permutation tests to calculate significance levels of yearly birth rate changes post-policy, identifying non-significant, transitory increases in 2016 and 2017
- Awarded a \$3,000 Charles Center Summer Research Grant; showcased findings at the 2022 W&M Fall Research Symposium