Zhongyi (James) Guo

A detail-oriented epidemiologist using data science and statistical skills to perform data analysis and model building.
(+1) 607-262-3415 ● guozy@stanford.edu ● Palo Alto, CA (willing to relocate) ● LinkedIn ● GitHub

Education

M.S., Stanford University, Palo Alto, CA 09/2023 – (Expected) 06/2025
Epidemiology & Clinical Research

B.S., Cornell University, Ithaca, NY 08/2020 – 05/2023
Double Major: **Biological Sciences** (Computational Biology) and **Biometry & Statistics** (Statistical Genetics)
Honors: **CUM LAUDE** (GPA: 3.57/4.3), **Dean's List**

Skills

- **Programming:** R, SAS, Python, Java, Swift, SQL, HTML, CSS, JavaScript, UNIX/Linux, LaTeX
- **Core Courses:** Advanced Epidemiology Methods, Biomedical Data Science, Data Fusion in medicine, Causal Inference, Data Mining & Machine Learning, Probability Models & Inference, Theory of Statistics, Population Genetics, Quantitative Genomics, Biostatistics, Object-Oriented Programming, Data Management using SAS
- **Core Skills:** Data Analysis and Visualization, Data Science, Machine Learning, Statistical Modeling and Testing, Microsoft Office, High-quality Multi-tasking, Highly Motivated & Supportive, Team Collaboration
- **Tools:** Git, GitHub, Jupyter Notebook, Xcode, RStudio, SAS Studio, Eclipse, Overleaf, Excel, Word, & PPT
- **Certification:** SAS Base (heading for SAS Advanced and CDISC)

Publication

Presenter/First Author, Causal effect of type II diabetes on prostate cancer in the East Asian population: A two-sample Mendelian randomization study, AACR Special Conference: Aging and Cancer, 2022 *(Published)*

Project Experience (Selected)

R package: hurdatPro (an unpublished class project) 05/2023
Co-author: He (Sarah) Zhang; GitHub Repository
- Created an R package in tar.gz format to analyze HURDAT Atlantic tropical cyclone activities with a teammate
- Cleaned data, wrote functions for storm plotting (track, position, size) and detecting landfall in US, computed accumulated cyclone energy of storms, and wrote unit tests for each function using testthat

Causal Effect of Type II Diabetes on Prostate Cancer in the East Asian Population 05/2022 – 12/2022
Advisor: Anqi Wang, Department of Preventive Medicine & Public Health Sciences, USC; GitHub Repository
- Performed two-sample Mendelian randomization with inverse variance weighted method on Type II Diabetes data from 23 East Asian GWAS and genetic-level GWAS summary statistics for prostate cancer from BioBank Japan
- Concluded that Type II diabetes has a negative causal effect on prostate cancer and presented at a poster session

GWAS Study: Analysis of Citrulline Levels and Chronic Kidney Disease 05/2022
Class Project; GitHub Repository
- Performed GWAS analysis on citrulline levels and chronic kidney disease data using two PCs obtained from Principal Component Analysis (PCA) as covariates on genotype data
- After reducing Type I error by Bonferroni correction, two significant SNPs were identified for two peaks on the Manhattan plot in R with two covariates added

α-helix or not? 12/2021
Class Project; GitHub Repository
- Performed feature engineering on the training set by averaging each feature of 5 neighboring amino acids and removed redundant features measured by correlation coefficient
- Trained binary classifiers (Logistic Regression, Decision Tree Regressor, and Random Forest models) using Sklearn in Python to predict α-helix or not using features derived from primary structures of proteins
- Tuned the maximum number of iterations using random search method to optimize the Logistic Regression model, and conducted cross-validation and examined the model accuracy (AUROC = 0.625)

Salaries in Big Techs 09/2021 – 12/2021

Class Project; GitHub Repository

- Built a multiple linear regression model to predict total yearly salaries based on employee features, including years of experience, gender, race, education, etc., for tech companies in US and overseas
- Established 3 equations for users to optimize their incomes by predicting total yearly salary in the US or overseas
- Detected gender and education inequity among different races in the US technology companies by modeling the relationship between [gender, race] and education levels using Logistic Regression

Research Experience (Selected)

Undergraduate Research Assistant at Basu Lab, Ithaca, NY 09/ 2022 – 12/2022

P.I.: Dr. Sumanta Basu, Department of Statistics and Data Science, Cornell University

- Applied Time Series Analysis in R to study time course and consequences of raised nocturnal blood pressure on cardiovascular disease among HIV-infected vs. uninfected adults in developing countries

Undergraduate Research Assistant at Lan Lab, Beijing, China 04/2021 – 07/2021

P.I.: Dr. Xun Lan, School of Medicine, Tsinghua University; Supervisor: Dr. Lihui Wang

- Designed a High Throughput Screening (HTS) to study the interaction between high-affinity TCR and pMHC (major histocompatibility complex) in human and used k-means clustering to create prototypes
- Found the optimal set of mutation locations in yeast to inhibit phosphoprotein functions by performing CRISPR/Cas9 every week and then analyzed the returned sequences of purified plasmids extracted

Teaching Experience

Cornell Bowers (College of Computing and Information Science)

- **Beta Tester & Teaching Assistant**, Introduction to Data Science 01/2023 – 05/2023
  - Graded homework, held office hours, proof-read assignments, & communicated between professor & students
- **Grader,** Probability Models and Inference 08/2022 – 12/2022

Teaching Assistant, Laboratory in Genetics and Genomics 01/2021 – 05/2021

- Created and stabilized knockout mutations on the target gene of fruit flies using CRISPR/Cas9
- Analyzed mutations vs. wildtype on the UCSC Genome Browser and piloted students to locate sgRNA transgene

Teaching Assistant (Summer), JNC Study Abroad Platform 07/2022 – 08/2022

- Two courses: Introductory Biology & Fundamentals of Physics I

Industry Experience

Match Group, Mobile (iOS) Development Intern (Remote) 06/2022 – 08/2022

- Replaced singletons using dependency injections in Match & Stir codebases using POP
- Developed a new feature named “Enhanced Interests” using SwiftUI that users can choose tags comprised of text and emoji under many categories to better deliver their interests to other people
- Replaced part of UIKit code with SwiftUI in watchOS and abridged AB Test to improve performance

Tencent, Data Analyst Project Intern (Remote) 07/2021 – 09/2021

- Web-scraped product details from e-commerce platforms & extracted data from databases using SQL
- Performed linear regression in Sklearn to detect customer preference patterns on different categories of products and practiced machine learning to simulate models to forecast the trend in item sales using Python programming
- Visualized and presented my analysis results to the marketing department for them to finalize marketing strategies

Extracurricular Experience (Selected)

Community HealthEd, Education Branch – Scientific Review Editor 03/2022 – 05/2023

- Performed literature review on each scientific paper/website cited in the articles to validate accuracy of citations
- Revised articles to remove technical jargon while retaining clear concise plain language to the general public as newsletters (materials primarily focused on maternal, prenatal, neurological & psychiatric health)
- Communicated efficiently between authors, copy editors, and the editor-in-chief