Zichen Liu

Los Angeles, CA | zichenliu@ucla.edu | [phone # hidden] | zichenliu.me

Summary

Hello! I am a biostatistics PhD student at UCLA. I have 3 years of industry experience in health data science & biomedical research. I enjoy programming and pick up new technical skills quickly, but my biggest strength is my effective scientific communication. I am looking for a summer internship in the biotechnology/biomedical industry.

Education

University of California, Los Angeles, PhD in Biostatistics (in-progress)	Sep 2022 – Present
University of Washington, MS in Biostatistics	Sep 2020 – Mar 2022
Pomona College, BA in Chemistry	Aug 2014 – May 2018

Highlights

Languages: R, Python, SQL, Julia, Java, Unix/Linux, Latex; some experience with Stata, SAS, Perl, C
Skills: Data science, data visualization, statistical modeling, machine learning (PyTorch, Keras, GLMnet, MCMC)
Other: Microsoft Office (Word, Excel, PowerPoint, Visio), agile project management (Asana, Jira), Git & GitHub

Experience

Graduate Student Researcher, UCLA Department of Biostatistics – Los Angeles, CA   Oct 2022 – Present
- Project 1: evaluating natural language processing algorithm for recruiting patients to ongoing cancer trial
- Project 2: developed R package to accompany upcoming textbook and designed figures for the textbook

Biostatistics Consultant, Public Health - Seattle & King County – Seattle, WA   Oct 2021 – Mar 2022
- Modeled housing instability after exiting public housing using survival analysis (incl. designing statistical analysis plans, presenting to stakeholders and housing partners, publishing article in housing policy journal)

Graduate Intern, Genentech – South San Francisco, CA   Jun 2021 – Sep 2021
- Modeled associations between respiratory disease outcomes and physical activity (incl. logistic regression, time-to-event analysis, analysis of variance) using high-dimensional wearable data from the UK Biobank

Data Specialist, Institute for Health Metrics and Evaluation – Seattle, WA   Aug 2020 – Feb 2021
- Built efficient Python pipeline to estimate US healthcare cost and utilization at the county/race level
- Designed a data warehouse to store incoming clinical data with automatic quality/redundancy checks
- Standardized and formatted 6 clinical datasets for the Global Burden of Disease Study using Python

Data Analyst, Institute for Health Metrics and Evaluation – Seattle, WA   Jun 2018 – July 2020
- Improved the disease modeling of 30+ injuries and sexual violence for the Global Burden of Disease Study with crosswalking to handle multiple data types and spatial-temporal smoothing methods
- Built artificial neural network in Python with Keras to detect new data sources from injury-related keywords
- Designed natural language processing algorithm to automate detection of ICD coding errors in extracted data

Publications

- Authorship on 17 publications (10k+ citations) in health policy and medical journals: bit.ly/zlpubs
- Presented at 3 national conferences: ACS 2018, APHA 2020, and the 2024 National LGBTQ Health Conference

Projects

powertools (R)   CRAN.R-project.org/package=powertools
- Developed 50+ functions for a variety of power & sample size calculations that solve for any parameter
- Designed rigorous parameter type/value checking suite to catch errors and print informative messages