319-855-3312
Seattle, WA
jinyutong0303@gmail.com

Yutong Jin

Profile
github.com/JinYuTong03
linkedin.com/in/yutong-jin-b3666a1b7/

Detail-oriented Masters-level statistician with expertise in R, Python, and SAS for healthcare data analysis and reporting. Skilled in supporting research initiatives both independently and in collaboration with multidisciplinary teams, managing large datasets, and preparing detailed research documentation. Experienced in healthcare research and passionate about advancing clinical outcomes through data-driven decisions.

Education

University of Washington, Seattle, WA
MS Biostatistics (Data Science) Sept 2023 - Mar 2025
University of Iowa, Iowa City, IA
BA Mathematics, BS Statistics (Mathematical Statistics) Jan 2020 - May 2023

Skills

Programming: R, Python, SQL, SAS, Git/GitHub, LaTeX, R Markdown, Microsoft Office, Power BI
Statistical Methods: Survival Analysis, Longitudinal Analysis, Bayesian Inference, Nonparametric Methods, Generalized Linear Models (Logistic Regression, Hypothesis Testing), Simulation, Sample Size Calculation, Machine Learning, Data Visualization
Clinical Data Expertise: Clinical Trial Design (Protocol Development, IRB Compliance), EHR/Patient Records Management, Data Validation & Quality Assurance (CDISC Standards), Categorical/Survey Data Analysis

Professional Experience

Fred Hutch Cancer Center, Paul T. Edlefsen Group, Seattle, WA
Research Assistant, Multi-Omics Integration Analysis Jan 2024 - Present
- Lead linear analyses of **multi-transcriptomics datasets**, applying Tilted Canonical Correlation Analysis to integrate scRNA-seq, identifying temporal transcriptomic shifts (HIV-1 LTR dynamics) and cell-type-specific enrichment patterns in B/T cells.
- Visualized co-embedding dynamics (common vs. distinct) using UMAP/heatmaps (ggplot2), prioritizing 2 biomarkers and 15 temporal genes via Spearman correlation.

Research Work

Kaiser Permanente Washington Health Research Institute, Yu-Ru Su's Group, Seattle, WA
Research Assistant, Evaluating Machine Learning Models for Selection Bias Adjustment in ACT Study Jul 2024 - Apr 2025
- Designed and implemented three machine learning models (**Lasso Regression, Random Forest, XGBoost**) to predict participant selection for autopsy, evaluating model performance with metrics like **calibration, AUC, and Brier score**.
- Conducted simulation studies and applied **inverse probability weighting (IPW)** techniques to address selection bias in the association analysis between neuropathology measures and Dementia, Alzheimer's disease.

Valley View Health, Seattle, WA
Retrospective Analysis of Multidisciplinary Diabetes Care Interventions Sept 2024 - Mar 2025
- Collaborated with Valley View Health Center to evaluate type 2 diabetes management interventions by cleaning and standardizing **38,000+ Electronic Health Records (EHRs)** from 11 sites, implementing robust data cleaning procedures, imputing **10 percent** missing data, and ensuring data quality for statistical analysis.
- Designed and implemented a **statistical analysis plan (SAP)** that incorporates **generalized estimation equations (GEE), linear regression models, and Cox proportional hazards regression** to assess longitudinal changes in A1C, blood pressure, and weight.
- Developed **data visualizations**, including **spaghetti plots, Kaplan-Meier survival curves and summary table**, to communicate key results effectively, improving stakeholder understanding and engagement.

Statistician, Examining Primary Care Integration in ACT Programs, Seattle, WA Mar 2024 - Jun 2024
- Led **protocol** design for a phase 3 multicenter **RCT** evaluating integrated primary care's impact on health outcomes in adults with serious mental illness (SMI), including standardized data collection, quality control, and management workflows.
- Developed comprehensive **statistical analysis plans** for randomized controlled trials, including randomization strategies, sample size calculations, and primary outcome analyses using the SF-36 health survey, ensuring methodological rigor and adherence to research standards.

WiDS Datathon: ADHD Diagnosis Prediction Jan 2025 – Apr 2025
Kaggle Competition Team Leader
- With colleague, developed ensemble model (XGBoost + Logistic Regression by Python) integrating fMRI connectome matrices with socio-demographic data, achieving 0.82 AUC in cross-validation.
- Implemented **Inverse Probability Weighting (IPW) in Python** to address selection bias and class imbalance, **improving model sensitivity by 15% for female ADHD detection**.
- Engineered **feature extraction pipeline using Python** for fMRI data with graph theory metrics (e.g., node degree, clustering coefficient), reducing dimensionality from **10k+ to 300 critical features**.