Bilin Nong

Mobile: (807)-706-7095

Personal Website
bilin.nong@mail.utoronto.ca

EDUCATION

University of Toronto (St. George Campus) 2021-09 – 2025-06
Honours Bachelor of Science - Statistical Science & Bioinformatics ASIP Co-op Program
 GPA: 3.96/4.00; Average: A+
 Undergraduate Courses: Multivariable Calculus, Linear Algebra, Probability & Statistics, Python Programming,
  Introduction to Databases, Data Analysis, Human Genetics, Molecular & Cell Biology

EXPERIENCE

Lunenfeld-Tanenbaum Research Institute @Sinai Health Toronto, ON
Research Assistant with Prof. Frederick Roth and Dr. Jochen Weile 2023-04 – Present
- Re-processed the raw data underlying existing variant effect maps (atlas of the functional effects of all possible genetic variants on proteins) with the latest versions of the TileSeqMave, a bioinformatics pipeline for the calculation of fitness scores from sequencing reads and provides a suite of quality control visualizations.
- Compiled benchmark sets of variants with known pathogenicity from online databases and literature for each map.
- Evaluated the predictions made by different versions of variant effect maps using the precision-recall curve and use them to infer evidence strength for clinical interpretation by calculating the log-likelihood ratio for pathogenicity.
- Gave recommendations for optimizing the implementation of TileseqMave pipelines based on analysis results.
- Delivered presentation on the methodologies and key finding of this project to the entire lab and wrote a report. (Link)

Dalla Lana School of Public Health @University of Toronto Toronto, ON
Research Trainee with Prof. Kuan Liu and Prof. Kevin Thorpe 2023-05 – Present
- Conducted a series of simulations in R statistical software to simulate Randomized Controlled Trials (RCTs) with different outcomes (continuous, binary, repeatedly measured) under different missing mechanisms.
- Applied common missing data handling methods such as complete case analysis, imputations and inverse probability weighting to each simulated data set, and assessed their performance by employing a range of evaluation metrics.
- Provided recommendations on how to deal with missing data in RCTs based on evaluation results.
- Acquired in-depth knowledge and expertise in missing data methodologies, as well as the design and analysis of RCTs, and gave a poster presentation on research showcase Day.(Link)

SKILLS

Technical Skills: SQL, Python, R, Unix Shell, Git, HTML, LaTeX
Libraries: psycopg2, tidyverse, doParallel, foreach

PROJECTS

Analysis of Sensory Factors & Coffee Ratings 2023-10 – 2023-12
- Investigated the correlation between the overall ratings of a cup of coffee and sensory aspects of coffee plus the bean species by fitting a multiple linear regression (MLR) model using the data from Coffee Quality Institute.
- Performed data analysis using R includes cleaning the raw data set, fitting the MLR model, checking the model assumptions via diagnostic plots, applying Box-Cox transformation to mitigate violated assumptions, conducting ANOVA and individual t test to check for significant linear relationship, and qualifying model goodness by a series of likelihood criteria.
- Utilized the tidyverse and car packages in R for efficient data summary, data analysis and visualization, showcasing proficiency in programming using R statistical software.
- Critically evaluated the model's limitations due to certain violations and the presence of problematic observations, demonstrating a thorough understanding of model reliability and data integrity.

Learning Management System Design 2023-09 – 2023-10
- Designed and implemented a schema for a learning management system database, specifically tailored to support the functionalities of MarkUs — a web application for the submission and grading of student assignments.
- Developed and executed complex SQL queries to facilitate data retrieval and analysis, demonstrating a deep understanding of relational database and SQL intricacies.
- Embedded SQL queries into Python using psycopg2 library, showcasing the ability to integrate SQL with a high-level programming language for efficient data handling and manipulation.
- Conducted testing and validation of database functionalities, ensuring accuracy and reliability of the data, and thereby facilitating insightful analytics for educational management and improvement.

Analysis of Hypertension & Low-Income Data in Toronto 2022-11 - 2022-12
- Implemented data cleaning and standardization processes, selected appropriate statistical methods for data analysis.
- Utilized the object-oriented programming paradigm to implement functions in Python and designed comprehensive test cases for each function to ensure validate program functionality.
- Analyzed and identified significant correlations between hypertension rates and low-income levels at the neighborhood level in Toronto, providing valuable insights into public health and socio-economic dynamics.
- Enhanced expertise in Python programming and data analysis libraries, demonstrating a strong foundation in data manipulation, statistical analysis, and result interpretation.

Honors and Awards

Summer Undergraduate Data Science (SUDS) Research Program Award 2023
Awarded to fund a summer research project for undergraduate students in data science field. ($7200 CAD)

Dean's List Scholar 2022, 2023
Awarded to students who had excellent GPA in the University