Xiaoling Yun M.S. in Biostatistics

4077 1 Richmond Street, New Brunswick, NJ, 08901 413-406-0188 Jacobyungt@gmail.com

SUMMARY

Data analyst, experienced in data management and biostatistics methods and various health data types: Medicaid claims, epidemiological, and genomic data. Proficient in complicated programming tasks in large datasets. Provide tables, graphs, technique manual and statistical analysis.

SKILLS

Machine Learning

Statistical Modeling and Interpretation

Data Management (SQL/Python/R/SAS)

Data Mining & Predictive Analytics

Data Visualization & Reporting(Power BI,/SAS/R)

Large Dataset Analysis

Technical Documentation & Communication

WORKING EXPERIENCE

Full time Data Analyst

Rutgers University Center for State Health Policy *2024 – Present*

- **Health Data Analysis**: Conduct statistical analysis and validation for New Jersey Medicaid cancer screening and sickle cell related research projects.
- **Medicaid & Claims Expertise**: Work with Medicaid claims and administrative data, including creating data formats, labels, and HEALTHCARE COST & UTILIZATION PROJECT relevant indicators updates and validation.
- **Statistical Programming**: Develop and implement statistical tests, tables, figures, and summaries using SAS, STAT and Power BI.
- **Homeless Information Management System**: program complex statistical tasks and validate results in collaboration with external universities.
- **Technical Documentation & Collaboration**: Review articles, prepare technique manuals for reproducibility, and work closely with investigators on research projects.

Research Assistant 05/2022 – 05/2023

University of Massachusetts Amherst, Ouyang Lab,

- Manipulated the ribosome-footprinting data (RNA-seq and Ribo-seq) using Pandas and Tidyverse, improving data processing efficiency.
- Analyzed translation efficiency and correlation of samples using RiboDiff, providing key insights into genetic expression.
- Visualized statistical results using advanced plotting techniques (heatmaply, ggplot2, and matplotlib) to support research findings.

Teaching Assistant 05/2022 – 05/2023

University of Massachusetts Amherst

Facilitated biostatistics education for 150 students in the "Introduction to Biostatistics" course, including hands-on instruction in STATA during lab sessions.

RESEARCH EXPERIENCE

Gene Expression-Based Acne Status Prediction: Uncovering a Signature Gene Set with Machine Learning 02/2023 – 05/2023

- Acquired gene expression data from GEO and preprocessed it by normalizing and scaling to ensure comparability
- Applied PCA to identify informative components and provide the visualization plot
- Employed LASSO to select relevant genes associated with acne by shrinking coefficients of less relevant genes
- Trained predictive models using SVM, LDA, and Random Forest algorithms, considering the expression levels of selected key genes

Conducted COVID-19 analysis using SAS 02/2022 - 05/2022

- Acquired and applied clinical trial programming concepts, including eCRF, SDTM, and ADaM, to generate domains and datasets (DM, DS, AE, ADSL, BDS, and OCCDS).
- Interpreted analysis plans and developed detailed Analysis Dataset Specifications to guide data processing and reporting.
- Executed data management tasks using SAS, including merging data sources and deriving variables with DATA steps, PROC SORT, and PROC SQL.
- Employed SAS functions (MACRO, PROC MEANS, PROC SGPLOT) to analyze data and create visualizations, facilitating comprehensive reporting.

Phase 2 Superiority Study Design for Migraine Treatment 02/2022 – 05/2022

- Defined primary efficacy and safety endpoints for the study, ensuring clinical relevance and regulatory compliance.
- Calculated sample sizes using statistical power analysis to optimize study design.
- Applied logistic regression models to determine the most effective dosage, contributing to successful study outcomes.

Analysis of 10 years of Chinese Stock A daily returns, focusing on validating the calendar effect 01/2021 - 05/2021

- Utilized financial service WIND to collect index data and employed Pandas for data cleaning
- Employed time series model to identify abnormal profit daily returns on a weekly basis
- Presented a stock strategy based on the identified effect

Python Data Mining for Stock Strategy Discovery 04/2020 - 07/2020

- Constructed a comprehensive stock database using SQL and ensured standardized data format by Pandas &Numpy.
- Developed a dynamic double moving average strategy with daily account updates
- Engineered an optimized buy-and-hold strategy for Chinese stock market portfolios, including white wine, pharmaceutical, and gold companies. Achieved an impressive 17.7% yearly return leveraging the Sharpe Ratio based on 2018-2020 data
- Discovered significant outperformance of beverage companies with over 100 yuan during Bull Markets, observed from 2006-2020 data

EDUCATION

University of Massachusetts Amherst — USA, 09/2021-05/2023

Master of Science in Biostatistics

• Shenzhen University — China, 09/2017-06/2021

Bachelor of Management in Accounting

Additional Information

Language: Chinese-native and English-fluent

Sports: 2019 Shenzhen University Economics College Basketball Championship, MVP