Yushan Gu

Linkedin: https://www.linkedin.com/in/guyushan/
Email: yushang@iastate.edu
Mobile: (801)541-7073

Education Background

- **Iowa State University** Ames, IA ***Ph.D.*** ***Statistics*** Aug 2017 - Jan 2024
- **University of Utah** Salt Lake City, Utah ***B.S.*** ***Mathematics*** (Dean's List) Jan 2013 - Dec 2016

Work Experiences

Data Scientist II, *Corteva Agriscience*
July 2023 - Present
Manager: Xiaoyi Sopko Indianapolis, IN

- Provided statistical expertise in the design, analysis, reporting, and interpretation of data related to the company's **seed** and **crop protection products**. Prepared relevant sections of **regulatory submissions, reports, and manuscripts**, which may include assessments of drug residues in soil, toxicity evaluations for organisms, or environmental impact analyses
  Effect: Critical for the market approval and maintaining registration of the company's products with regulatory authorities
- Collaborated with cross-functional and interdisciplinary teams to solve problems, perform statistical modeling/data analysis, and interpret complex datasets. Developed customized tools, such as RShiny applications, to generate specific results as needed

Data Scientist Co Op, *Bayer Crop Science*
Jan 2023 - Jun 2023
Manager: Prabhakar Dhungana Remote

- **Predicted the flowering time** in male corn by utilizing a **mixed-effects model** and **Growing Degree Units (GDUs)**
  Effect: Aid researchers and farmers in crossbreeding corn with greater precision. This model is also applicable for predicting other critical periods, such as the optimal harvesting time, potentially reducing labor and warehousing costs. The approach can be similarly adapted for other crops
- Independently developed algorithms in **R**, **Python** and **SQL** to clean, process, analyze and visualize multivariate datasets
- Worked closely with data science team to break down complex problems to build feasible solutions that are impactful, accurate and validated

Research Assistant, *Iowa State University*
Jul 2019 - Jan 2024
Advisor: Yumou Qiu

- **Developed a novel method** that extends and combines the **Lasso** method and **MANOVA**, allowing simultaneous inference of multivariate coefficients in the **High-Dimensional** Linear Regression Model with either univariate or **correlated multivariate responses**. The proposed method achieved a **more accurate** Type I error than the existing multivariate testing methods
- Used the proposed method to model and analyzed data from different sources:
  - **Agricultural Data** – Investigated the significance of the effect of certain phenotypes on the Ear, Cob, and Kernel of plants
  - **Genetic Data** – Identified a group of significant Single Nucleotide Polymorphisms (SNPs) related to specific phenotypes. Our method has **more stable results** than FarmCPU, which is a Genome-Wide Association Study (GWAS) method
  - **Stock Data** – Analyzed the stock price data from S&P 500 in 2008. Studied the impact of other stocks on the return of JPMorgan Chase (JPM). The significant industry groups identified by our method are **more explainable** than the de-correlated score approach (DCScore) and the de-correlated score approach (LP)

Data Scientist Intern., *Chongqing Zhida Education Technology Co., Ltd.*
Jul 2021 - Mar 2022

- Used **SQL** to store, retrieve, and manipulate data in databases; Statistically analyzed data in **R**
- Applied **unsupervised Machine Learning** techniques (E.g., **Factor Analysis**, **Principal Component Analysis (PCA)**, and **K-Means Clustering**) to classify students into several groups; Applied different treatments on groups to develop the optimal plan for improving student achievement
- Applied **supervised Machine Learning** technique (E.g., **Classification Tree**) to train a model based on the survey results from students in different majors, to advise students who are still on the fence
- **Visualized** student grades so that parents and teachers can better track and adjust teaching methods and directions

Research Experiences

A Predictive Model for Success Rate of Jejunal Feeding Tube Placement
Jan 2022 - May 2022
Collaborate with: Hui Gao

- **Predicted** the success rate of tube placement with patients' features by different models, including **Logistic Regression Model**, **Classification and Regression Trees (CART)**; Reduced overfitting by **Cross-Validation**
- Compared the performances of different models by **Calibration Plot**, **Decision Curve**, and **AUC**
- Used **ROC Curve**; Found the best cutoff point; Built prediction model with relatively **low prediction error**
- **Visualized** the models in **Nomogram**

Technical Strengths

- **Programming:** R, SQL, Java, Python, SAS **Other Skills:** RShiny, JavaScript, HTML, Git, SPSS, Excel VBA, Photoshop
- **Quantitative Skills:** Multivariate ANOVA (MANOVA), Generalized Linear Mixed-Effects Model, Survival Analysis, Machine Learning (Spectral Clustering, Random Forest, Neural Network. . .), Classification, Missing Data, Special Data, Bayesian Statistics, Data Modeling, Data Visualization, Data Structure, Algorithms

Selected Publication

- Qiu Y, **Gu Y**. 2024. F Statistics for High-Dimensional Inference of Linear Model. *Bernoulli*. (Accepted)
- Hung M, Smith WA, Voss MW, Franklin JD, **Gu Y**, Bounsanga J. 2019. Exploring student achievement gaps in school districts across the United States. *Education and Urban Society.*
- Hung M, Voss MW, Bounsanga J, **Gu Y**, Granger EK, Tashjian RZ. 2018. Psychometrics of the Patient-Reported Outcomes Measurement Information System Physical Function instrument administered by computerized adaptive testing and the disabilities of arm, shoulder and hand in the orthopedic elbow patient population. *Journal of shoulder and elbow surgery* 27 (3), 515-522.