Zehan Yang

Github: https://github.com/YEnthalpy
LinkedIn: https://www.linkedin.com/in/Zehan

zehan.yang@uconn.edu
860-634-0053

QUALIFICATION

- Strong background using R, SAS and Python for research and projects.
- Sufficient experience in dealing with medical data.
- Proficient in statistical modeling using general linear model.
- Proven ability to identify, analyze, and solve problems.
- Worked both as a member and independently.

EDUCATION

- **University of Connecticut** Storrs, CT, USA
  Ph.D. in Statistics; GPA: 3.93/4.3 *Aug 2023(Expected)*

  Relevant course: Statistical Computing, Multivariate Statistics, Design of Experiments, Applied Statistics, Mathematical Statistics, Linear Statistical Models, Statistical Inference, Survival Analysis.
  Research Interest: Optimal Subsampling for Big Survival Data.

- **Xi'an Jiaotong University** Xi'an, China
  Bachelor of Computational Mathematics; GPA: 3.43/4.3 *May 2018*

SKILLS SUMMARY

- **Languages**: R, SAS, C++, Matlab, Python
- **Tools**: LaTex, Git, HPC Cluster, Slurm Workload Manager

EXPERIENCE

- **University of Connecticut** Storrs, CT, USA
  Research Assistant - Prof. Jun Yan & HaiYing Wang *April 2020 - Current*

  - **Optimal Subsampling of Parametric Accelerate Failure Time (AFT) model for Big Survival Data**:
    - Proved the asymptotic properties of the resultant estimator in the general subsampling approach.
    - Proposed optimal subsampling probabilities based on A-optimality and L-optimality from the optimal design of experiment.
    - Developed the two-step algorithm to get the resultant estimator using the optimal subsampling probabilities.
    - Proved the asymptotic properties for the estimator derived by the two-step algorithm and propose the formula to estimate the variance and covariance matrix.
    - Applied the simulation and the real data analysis based on parametric Weibull AFT model.
  - **Optimal Subsampling of Semi-Parametric Accelerate Failure Time model for Big Survival Data**:
    - Proposed a general estimating function for semi-parametric AFT model based on the least-squares approach .
    - Used resampling method to construct a sandwich estimator for the variance-covariance matrix.
    - Constructed A-optimal subsampling probabilities by the slope matrix that was the 'bread' part of the sandwich estimator.
    - Developed the two-step algorithm and the iterative algorithm to get the resultant estimator using the optimal subsampling probabilities.
    - Compared the computational and estimation efficiency for two algorithms.
    - Applied the simulation based on different error distributions. Finished the real data analysis based on a big lymphoma data.

- **University of Connecticut, Health Center** Farmington, CT, USA
  Research Assistant - Prof. Helen(Zhao) Wu *Sep 2021 - Current*

  - **Impact of Mental Disorder and Drug Addiction on Adult Bio-markers Among Low-income Women after Hurricane Ike**:
    - Cleaned the Hurricane Ike data for 1000 observations.
    - Selected patients who had bio-marker testing results before and after Hurricane Ike.
    - Used an ANOVA model to investigate the influence of age, ethnicity, marital status, employment status and annual income on the changes in biomarkers.
    - Implemented a mixed ANOVA model to detect if patients' bio-markers, such as 'IL6', 'Cortisone', 'DHEAS', changed after experiencing Hurricane Ike for different kinds of drug users and mentally disordered patients.
  - **Multiple Consulting projects for medical researchers and medical doctors**:
    - Applied a logistic regression to analyze the relationship between breast cancer biopsy status and patients' personal information based on breast cancer data in Sudan.

- Analyzed Medication Data for diabetics to evaluate if the CPAP machine helped patients control their A1C score.
- Applied Beta regression models to analyze the National Health Service (NHS) Survey data.

- **University of Connecticut** Storrs, CT, USA
  Independent Teaching Assistant *Sep 2021 - Current*
  - **STAT 3375Q Introduction to Mathematical Statistics I Teaching Assistant**:
    Lead discussion sessions and hold office hours weekly.
  - **STAT 3445 Introduction to Mathematical Statistics II Teaching Assistant**:
    Lead discussion sessions and hold an office hour weekly.
  - **STAT 1000Q/1100Q Introduction to Statistics Teaching Assistant**:
    Teaching students how to use Minitab during the discussion and grading their Minitab assignments weekly.

PROFESSIONAL ASSOCIATIONS

- Membership of American Statistical Association
- Membership of New England Statistical Society

PRESENTATIONS

- Optimal Subsampling for Parametric Accelerated Failure Time Models with Massive Survival Data. *2022 Joint Statistical Meetings*, Washington D.C., August 6 - August 11, 2022.

HONORS AND AWARDS

- The 2022 LiDS Student Paper Award by the American Statistical Association.

PUBLICATIONS

- Yang, Z., H. Wang, and J. Yan (2022). Optimal Subsampling for Parametric Accelerated Failure Time Models with Massive Survival Data. *Statistics in Medicine 41(27)*, 5421–5431.