

# Yiling Huang

yilingh@umich.edu - (734) 450-6430

## EDUCATION

### University of Michigan-Ann Arbor

*Ph.D. in Statistics*

Ann Arbor, MI

*Expected May 2027*

- GPA: N/A

- Research interests: statistical learning; selective inference; causal inference; interpretable learning algorithms

### University of Michigan-Ann Arbor

*B.S. with Highest Honors in Statistics; B.S. in Mathematics (Pure Math Program)*

Ann Arbor, MI

*May 2022*

- GPA: 3.98, James B. Angell Scholar & University Honors

- Honors Thesis Title: “Balance Assessment of Matched Data with Multiple Treatment Levels” [Link](#)

### University of Rochester

*First-year Student with Focus in Statistics*

Rochester, NY

*September 2018 – May 2019*

- GPA: 4.00, Dean’s List

## HONORS & AWARDS

*Best Undergrad. Presentation, Michigan Student Symposium for Interdisciplinary Statistical Sciences*

*March 2022*

*Outstanding Achievement in Mathematics Award, Dep. of Mathematics, Univ. of Michigan*

*April 2022*

## RESEARCH EXPERIENCES

### Department of Statistics, University of Michigan

Ann Arbor, MI

*Undergraduate Researcher (Supervised by Prof. Snigdha Panigrahi)*

*May 2021 – Present*

- (*In progress*) provide post-selection inference for sparse graphical models applicable to Gaussian data, and reviewed literature on recent advances in selective inference
- Derived characterizations of the randomized paired group lasso’s model selection event, and extended the results to derive the likelihood of the statistical pivot conditioning on the model selection event
- Implemented a randomized paired group lasso algorithm that estimates sparse precision matrices of multivariate Gaussian data

### Department of Statistics, University of Michigan

Ann Arbor, MI

*Undergraduate Researcher & Honors Thesis (Supervised by Dr. Mark Fredrickson)*

*January 2021 – Present*

- Developed a hypothesis test that assesses covariate balance of matched (grouped) datasets with multilevel treatments, especially observational data that are matched using propensity score-based methods
- Adopted the adjacent category logistic regression model to represent treatment assignment probabilities, and tested the validity of matching results by testing the equivalent formulation that all slope parameters equal zero
- Derived exact expressions of the test statistic after accounting for stratification information via conditioning on strata-based statistics, and conducted simulations to evaluate operating characteristics of the test

### Department of Statistics, University of Michigan

Ann Arbor, MI

*Undergraduate Researcher (Supervised by Prof. Kerby Shedden) [Link to Report](#)*

*September 2021 – Present*

- Provided statistical methods for the analysis of time-series, dyadic clinical data by designing an optimization problem that aims to combine the strength of PCA and CCA
- Implemented several gradient-based optimization algorithms, including BFGS, to solve the optimization problem
- Experimented on simulated data to confirm the proposed method successfully reconciles PCA with CCA

### Yang Zhang Lab, University of Michigan

Ann Arbor, MI

*Research Assistant*

*March 2020 – May 2021*

- Implemented optimization algorithms including differential evolution (DE) and grid search to optimize parameters of a bioinformatics predictive model that predicts protein structures
- Extracted features from protein sequential data with C++ programs
- Studied multiple bioinformatics journal papers that perform protein structure prediction using deep learning and optimization algorithms, and proposed feasible improvements or alternatives as future research directions

## PROJECTS

### Regression Analysis (STATS600, Ph.D. level, individual work) [Link to Report](#)

- Conducted regression analysis on high-dimensional human gene expression level data to detect genes highly associated with cancer severity; deployed the LASSO for variable selection
- Proposed a method to obtain an *a posteriori* modeling assumption for linear regression models; performed simulations to conclude that the method yields confidence intervals of proper length and coverage
- Applied variable selection and statistical inference separately using two-fold sample-splitting to avoid invalid post-selection inference

## PROJECTS (CONTINUED)

---

### **Statistical Learning (STATS601, Ph.D. level, group work)** [Link to Report](#)

- Implemented statistical/machine learning algorithms to make predictions about whether the card owners will abandon a specific credit card service based on their consumption and demographic features
- Compared the performance of linear models, tree-based models, and neural nets on the credit card dataset
- Contributed the idea of training learning algorithms separately by cluster to learn heterogeneous patterns on each cluster of data, and the writing and implementation of relevant parts (sections 1-3, 6-8, 10, and 11) of the report

### **Computational Methods in Statistics and Data Science (STATS406, individual work)** [Link to Report](#)

- Implemented density estimation algorithms to estimate the total variance distance (TVD), proposed the usage of this quantity as a test statistic for permutation tests, and compared its performance to other statistics
- Performed permutation tests to quantify the distributional differences between the abilities of soccer players from different countries
- Generated bootstrapped confidence intervals for the parameters of a random effects linear model that aims to explain player abilities by individual-level and country-level covariates

## TALKS & PRESENTATIONS

---

### **Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS)**

March 2021

- Balance Assessment of Matched Data with Multiple Treatment Levels ([link to poster](#))

### **American Causal Inference Conference (ACIC)**

May 2021

- Balance Assessment of Matched Data with Multiple Treatment Levels ([link to poster](#))

## WORK EXPERIENCE

---

### **Department of Mathematics, University of Michigan**

Ann Arbor, MI

#### *Grader for Advanced Calculus I*

September 2021 – Present

- Graded homework for two sections with a total of 50 students and produced answers to all problem sets
- Selected a set of proof-based problems from the instructor's assigned lists as homework problems
- Designed the grading scale for each problem based on the specific structure of valid proofs

## COURSEWORK

---

### **Graduate Courses at the University of Michigan:**

- Regression Analysis (A; Ph.D. level), Statistical Learning (A; Ph.D. level), Linear Programming (A; master level), Statistical Inference (A; master level)

### **Undergraduate Courses at the University of Michigan:**

- Linear Algebra (A), Advanced Calculus I (A+; mathematical analysis), Probability Theory (A+, weakly measure-theoretic), Introduction to Modern Algebra (A+), ODE (A), Point-set Topology (A+), Real Analysis (A+), Combinatorics (A+)
- Introduction to Theoretical Statistics (A+), Data Mining and Statistical Learning (A+), Bayesian Data Analysis (A), Applied Regression Analysis (A+), Computational Methods in Statistics and Data Science (A)
- Data Structure (A; in C++)

### **Undergraduate Courses at the University of Rochester:**

- Multivariable Calculus (A), Introduction to Probability (A), Discrete Mathematics (A)
- Introduction to Categorical Data Analysis (A)

## SKILLS

---

- Languages: Chinese (native), English (fluent)
- Technology: Data analysis, processing, and visualization in R and Python; OOP programming in Java, C++, and Python; MATLAB programming; typesetting in LaTeX; Final Cut Pro
- Leadership: Captain of the delegation team of Capital Normal University High School (CNU-HS) for BGBC, one of Beijing's citywide high school soccer championships in 2018