

# ALEX ZHAO

**Research interests:** high-dimensional inference, classification and regression, natural language processing

331A Thomas Building, University Park, PA 16801  $\diamond$  yazhao@psu.edu

## EDUCATION

---

### **Pennsylvania State University**

Doctor of Philosophy, Statistics

Advisor: Dr. Runze Li

### **University of Chicago**

B.A. in Economics, Minor in Statistics

Dean's List (2008-2012)

June 2012

## RESEARCH AND PROFESSIONAL EXPERIENCE

---

### **Pennsylvania State University**

*Graduate Assistant, Department of Statistics*

August 2016 - Present

*University Park, PA*

- Currently developing new methods for conducting hypothesis tests for high-dimensional regression coefficients, with applications to linear regression and GLM settings

### **Software Engineering Institute**

*Science of Cybersecurity Group Intern, CERT Division*

May 2017 - August 2017

*Pittsburgh, PA*

- Worked on enhancing gradient boosting classification algorithm for Natural Language Processing (NLP) project
- Added functionality to and maintained internal R package, built out reporting tools for measuring model performance

### **Zynga**

*Senior Marketing Data Analyst*

January 2015 - July 2016

*San Francisco, CA*

- Fit statistical and machine learning models to project lifetime value and behavior for Zynga's mobile game players
- Designed data mining scripts and queries for creating custom audiences and tracking for user acquisition campaigns

### **GrubHub**

*Marketing Analyst*

October 2012 - January 2015

*Chicago, IL*

- Created likelihood models for repeat user behavior and regression models for TV advertising
- Taught SQL and relational database concepts to Marketing and Strategic Development team members
- Created reporting and compiled research for marketing and PR using Tableau, SQL, Python, and other tools

## TEACHING EXPERIENCE

---

### **Graduate Student Online Teaching Certificate, Pennsylvania State University**

Spring 2020

### **Pennsylvania State University, Department of Statistics**

*Instructor*

Fall 2016 - Present

*University Park, PA*

STAT 319: Applied Statistics in Sciences - Fall 2019, Spring 2020

STAT 415: Introduction to Mathematical Statistics - Summer 2019

*Lab Instructor*

STAT 200: Elementary Statistics - Fall 2016

*Guest Lecturer*

STAT 470W: Problem Solving and Communication in Applied Statistics - Fall 2020, "Introduction to SQL"

STAT 380: Introduction to Mathematical Statistics - Spring 2018, "Web Scraping"

### **University of Chicago, College Core Tutors**

*Economics Tutor*

Fall 2011 - Spring 2012

*Chicago, IL*

Elements of Economic Analysis I-IV (ECON 200-203)

Introduction to Microeconomics (ECON 198)

Introduction to Macroeconomics (ECON 199)

## SERVICE AND ACTIVITIES

---

<b>Penn State Graduate and Professional Student Association</b>	President (2020-Present) Treasurer (2018-2020) Assembly Delegate, College of Science (2017-2018)
<b>Statistics Graduate Student Association</b>	President (2019-2020) Vice President (2017-2019) Treasurer (2016-2017)
<b>Depaul Debate Society</b> <b>Chicago Debate Society</b>	Head Coach (2012-2015) Top 16, 2012 APDA National Championship APDA League Expansion Coordinator (2011-2012)

## TECHNICAL SKILLS

---

<b>Research</b>	R, LaTeX, Python, SQL, Tableau, Matlab, Stata, Minitab
<b>Other</b>	Excel, Google Analytics, HTML, WordPress
<b>Languages</b>	Mandarin Chinese (native speaking fluency), Spanish (working knowledge)

## HONORS AND AWARDS

---

**Student Leadership Scholarship (2020-2021)**  
**Eberly College of Science Dean's Climate & Diversity Award Nominee (2020)**  
**Mu Sigma Rho Honor Society (2019)**

## PUBLICATIONS

---

**Zhao, A., Li, C., Li, R.** "Testing Regression Coefficients in a High Dimensional Empirical Likelihood Setting." In Prep.

## OUTSIDE RESEARCH PROJECTS

---

<b>Penn State OIS Account Matching Algorithm</b>	2019 - 2020
Created machine learning models to allow the Office of Information Security better identify duplicate user accounts that were created in error to replace existing heuristic algorithms.	
<b>Text-Based Classification of Technical Debt</b>	Summer 2017
Project to build machine learning algorithm for identifying software technical debt. Improved model's use of n-grams, incorporated non-text information, and created pipeline for evaluating model performance and iterations.	

## PRESENTATIONS

---

<b>Penn State Statistics Graduate Student Association Workshop, Speaker</b>	Spring 2021
How to Make a Website	
<b>Penn State Statistics Graduate Student Association Workshop, Speaker</b>	Fall 2020
Introduction to LaTeX	
<b>Penn State Statistics Graduate Student Association Workshop, Speaker</b>	Spring 2020
How to Make a Website	
<b>Penn State Statistics Graduate Student Association Workshop, Speaker</b>	Fall 2019
Introduction to LaTeX	
<b>Penn State Libraries R Users Lightning Talk, Speaker</b>	Fall 2018
The quanteda R Package	
<b>Penn State Statistics Graduate Student Association Workshop, Speaker</b>	Fall 2018
Introduction to LaTeX	

## POSTERS

---

<b>2020 Joint Statistical Meetings, Philadelphia PA (Virtual)</b>	Summer 2020
Extending a Large Dimensional Empirical Likelihood Test to Regression	

## PROFESSIONAL DEVELOPMENT

---

<b>MOPTA Summer School on "Optimization and Learning"</b>	Summer 2018
Attended summer workshop on Pytorch, online learning and stochastic gradient descent, and nonconvex optimization methods	