

# RYAN ZHANG

609-751-7639 | rczhang@andrew.cmu.edu | <https://rzryanz00.github.io>

## EDUCATION

**Carnegie Mellon University, GPA: 4.00/4.00**

*Bachelor of Science in Statistics and Machine Learning*

Pittsburgh, PA

*Expected May 2025*

## PROFESSIONAL EXPERIENCE

**Regeneron Pharmaceuticals**

May 2024 – Aug. 2024

*Artificial Intelligence and Machine Learning Intern*

- Applied statistical power analysis to determine the optimal sample size for a \$7 million commercial pilot study; authored and presented a detailed report to executive leadership
- Developed a Text-to-SQL application leveraging Large Language Models (LLMs) from Hugging Face, enabling the conversion of natural language inputs into SQL queries to aid data retrieval
- Applied an accelerated failure time model utilizing XGBoost to forecast 300,000 patient journeys across various lines of therapy
- Implemented a GARCH time series model to forecast the variance of financial payments from pharmaceutical companies to physicians
- Compared predictive performance of transformer models, N-BEATS models, and BlockRNN models with 2 years of pricing and volume time series data

**Regeneron Pharmaceuticals**

May 2023 – Aug 2023

*Data Science Intern*

- Leveraged statistical models like Kruskal-Wallis and Levene's test to analyze 10 million+ sampling data, optimizing quality control sampling for biopharmaceutical clean rooms, resulting in a 33% increase in sampling efficiency
- Streamlined data processing with regular expressions and the pandas library in Python, employing vectorization for enhanced computational efficiency, and identified 50+ inconsistencies throughout the examination process
- Designed and implemented a team-wide program to automate error identification, ensuring data accuracy and compliance with regulatory standards

## COMPETITIONS

**Optiver Market Making Competition**

Mar. 2024

*3rd Place (Team of 5)*

- Approximated the value of contracts, which were determined by a combination of math, probability, brain teasers, and estimation questions
- Traded contracts to capitalize on other teams' mis-pricings

**Cornell Trading Competition**

Oct. 2023

*1st Place (Team of 4)*

- Selected from 775 intercollegiate undergraduate and graduate applicants to compete as a team of four in three algorithmic case studies: systematic equities, derivatives, and crypto
- Executed an algorithmic market-neutral approach by employing cross-momentum strategies to drive profits and validated results through rigorous back testing
- Presented findings and winning strategies to an audience of 40+ intercollegiate teams and event directors

## PROJECTS

**Tabular Magic | Python**

May 2024 - Present

- Heavily contributed to a Python package designed for low-code data science, streamlining processes like dataset exploration, statistical and regression analyses, and machine learning model benchmarking
- Designed for clinical researchers with limited statistical backgrounds
- Currently co-authoring a manuscript focused on clinical data

**COVID Hospitalization and Wastewater Forecasting | R**

Aug. 2024 - Present

- Using time series analysis to evaluate whether incorporating wastewater data enhances the accuracy of COVID hospitalization forecasting
- Employed autoregressive quantile regression as the primary modeling approach

**Soybean Futures Time Series Forecasting | R**

April 2024

- Forecasted soybean futures based on weather data from Quincy, Illinois.
- Compared multivariate and univariate time series models such as vector autoregression and GARCH.

## TEACHING

---

### **36-236: Probability and Statistical Inference II**

Jan. 2024 – May 2024

- Led weekly recitations, hosted office hours, graded exams and homework

### **36-235: Probability and Statistical Inference I**

Aug. 2023 – Dec. 2023

- Led multiple weekly recitations, graded exams and homework

### **36-226: Introduction to Statistical Inference**

Jan. 2023 – May 2023

- Hosted office hours, graded exams and homework

## SKILLS

---

**Computer Science:** Python (NumPy, pandas, PyTorch, scikit-learn), C, R, SML, Tableau, Plotly Dash