Department of Statistical Sciences (DoSS) 2026 Summer Undergraduate Research Awards

The Department of Statistical Sciences is seeking applicants interested in conducting a summer research project under faculty supervision. Projects will take place for a 16-week period from May – Aug 2026. The following research awards are being offered this summer: NSERC USRAs ($7,500), UTEA ($7,500), DoSS SURAs ($2000).

All students who are currently enrolled in a Department of Statistical Sciences (DoSS) Specialist or Major program and who have completed STA302H1 with at least a B+ by the time of application are eligible to be considered for one of these awards. You will be asked to rank your top 3 projects. Students will be considered based on their academic record, experience, and statement of research interest. Students who are shortlisted for specific projects may be invited for an interview with the prospective supervisor.

Projects

Analysis of internet data

Supervised by: Radu Craiu

The project consists of data scraping (gathering data from internet repositories) and analysis. The area of interest will be discussed with the student. The methods will rely heavily on programming skills. The statistical methods used are likely to cover a large range so some adaptive learning will also be needed.

Exploring complexity finance

Supervised by: Leonard Wong

Mainstream financial economics and mathematical finance focus on equilibrium which is theoretically more tractable. An alternative approach, based on the theory of complexity systems, allows disequilibrium and diverse emergent behaviours. In this project, we will explore complex systems in the financial context and implement several models, including agent-based simulation of financial markets.

A 25-Year Review of Statistics Education Research in Canada

Supervised by: Bethany White

Scholarship of Teaching and Learning (SoTL) is the systematic inquiry of student learning, and the sharing of results. This research project will involve a literature review of Statistics-focused SoTL published in peer-reviewed journals between 2000 and 2025 by authors affiliated with Canadian postsecondary institutions. The goals of this work are to describe the body of work produced in this area, identify pedagogical and methodological themes in the literature, and examine how these patterns have evolved over time. By mapping publication trends, research designs, instructional settings, and author characteristics, the study will offer insights on Statistics SoTL activity in Canada over the past 25 years.

Motion correction in fMRI analysis

Supervised by: Jun Young Park

Functional magnetic resonance imaging (fMRI) is a widely used tool for studying how brain signals, measured as time series, are associated with cognition and behavior. During an fMRI scan, participants are instructed to remain still, because head motion can substantially degrade data quality. When motion is severe, the resulting data may become unusable for scientific analysis. To control these unwanted, non-biological, and artifactual variations, several approaches have been developed to reduce motion effects in fMRI data, including (i) data scrubbing, (ii) statistical modeling, and (iii) deep learning–based methods. Each approach has its own advantages and limitations, and there is currently no gold standard. Moreover, these methods have rarely been evaluated in a comprehensive and systematic way from a statistical perspective. In this project, we will work with large, real-world neuroimaging datasets to investigate how different motion-correction methods behave in practice, and to identify situations in which one approach is expected to perform better than others. The student will gain hands-on experience in data analysis, statistical reasoning, and scientific computing, while contributing to an ongoing research project in neuroimaging and data science.

Quantifying Environmental Action in Faith Communities

Supervised by: Meredith Franklin

Religion is frequently cited as a powerful driver of environmentalism, with the capacity to shape values, influence behavior, and mobilize communities. Despite this potential, there is a lack of comparable, empirical evidence about the scope, distribution, and determinants of faith-based environmental action across traditions and regions.
This project addresses this gap through the development of the Green Faith Database Initiative, a scalable, continually updatable database of environmental actions taken by places of worship using responsible web scraping, text analytics, and human-guided Large Language Models (LLM) labelling with validation metrics and verified confidence scores.

Computational Methods for Detecting Major Agents in Mean Field Game Models

Supervised by: Ricardo Baptista and Dena Firoozi

Many large-scale systems, including financial markets, power grids, and opinion dynamics, exhibit hierarchical interactions in which a small number of influential agents significantly affect the behavior of a large population. Models for mean field games with major and minor agents provide a framework for capturing such asymmetries. However, existing formulations typically presume that the identity of a major agent is known a priori, while in practice the major agent may be only indirectly identifiable. This project will develop statistical tools to infer the existence and role of a major agent from observed data. In particular, we will use likelihood-based statistical tests to identify the major agents in finite-population mean field games. We first consider a linear–quadratic–Gaussian setting with one major agent and a large but finite number of minor agents whose dynamics are coupled through the population's average state (empirical mean field). Theoretical results will be complemented by numerical studies on synthetic data to validate performance under controlled conditions, as well as real-world datasets where major agents are suspected to exist. Students will learn statistical computation and have the opportunity to explore stochastic control, mean field game theory, and statistical inference in this summer project.

Fusing satellite images with high-frequency ground monitoring for modeling and detecting methane plume events

Supervised by: Meredith Franklin

Methane emitted from oil and gas production is a major anthropogenic driver of climate change. This project will link methane satellite images to a stationary ground sensor with high temporal resolution to quantify the spatial and temporal patterns in methane emissions over a major oil and gas production area of the US. Statistical and machine learning modeling will be used with the fused data to detect methane plumes (i.e. extreme methane events from gas leaks or operational emissions). Several methodological approaches will be tested and compared including quantile regression and Gaussian process modeling. Sensitivity analyses, probability calibration, and uncertainty quantification are additional components that will be considered in the workflow.

Developing spectrograms for high-dimensional sound data to classify fish species

Supervised by: Vianey Leos Barajas

Traditionally, fish population abundances are tracked through invasive capture methods which require time, labour, and material investments and result in the mortality of many fishes. Hydroacoustic surveying has become an alternative to invasive capture methodologies and is currently being tested by the Ontario Ministry of Natural Resources as a possible alternative approach. In this, sonar is used to locate organisms and objects in the water and the sound emitted at up to 400 distinct frequencies bounces off organisms back to a receiver. These signals received may act as a species "fingerprint" allowing the classification of species and abundance calculations.

In this project, students will develop spectrograms for high dimensional data for use in classification of fish species. Students will also be tasked with applying various machine learning and AI tools to spectrograms.

K-means for Processes

Supervised by: Sebastian Jaimungal

The project is built around a barycentre-based notion of clustering, where each stochastic process is identified with its law and cluster centres are defined as barycentres in a space of probability measures on stochastic processes. These barycentres are characterized as solutions to variational problems whose optimality conditions lead to nonlinear PDEs. The student will study how these barycentres can be computed by evolving distributions through controlled diffusion dynamics, providing a natural link to diffusion models in modern machine learning. In particular, the project will draw parallels between barycentre computation and score-based diffusion models, where neural networks are trained to approximate drift or score functions. This perspective unifies stochastic control and diffusion-based generative modeling and allows the student to explore how ideas from machine learning can be used to efficiently compute barycentres of stochastic processes.

Student Eligibility:

1. Must be an undergraduate student currently enrolled in a Department of Statistical Sciences Specialist or Major program.
2. Currently registered full-time or part-time student at the time of application.
3. Must have completed STA302 with at least a B+ grade.

How to submit your application:

Please fill out and submit the following application form:

Application for DoSS Summer Undergraduate Research Awards 2025

If you have any questions regarding these awards, please contact ug.statistics@utoronto.ca

For more information, please visit: Statistical Sciences Research Opportunities and Awards

Completed applications are due by **Thursday, February 12th, 2026, at 11:59PM**