Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

Project code:	SCI028
Project title:	**Can AI Learn to Queue? Learning Routing Decisions in Service Systems**
Discipline:	Department of Statistics
Supervisor(s)	Azam Asanjarani
Contact details	azam.asanjarani@auckland.ac.nz
Skills Needed	• Basic programming skills • Interest in data science, machine learning, or artificial intelligence • Strong analytical and problem-solving skills • Basic knowledge of probability and statistics • Ability to work independently and communicate findings clearly

Project description

Queueing systems arise in many real-world applications, including call centres, hospitals, cloud computing platforms, and transportation systems. A fundamental challenge is determining how arriving jobs should be assigned to available servers. Classical routing policies, such as Join-the-Shortest-Queue (JSQ), have been extensively studied and are known to perform well under many conditions.

This project investigates whether machine learning models can learn routing decisions directly from data generated by queueing simulations. The student will develop a simulation of a multi-server queueing system and generate data under established routing policies. Using this data, several machine learning methods, such as logistic regression, decision trees, and random forests, will be trained to predict routing decisions based on the current state of the system.

The project will compare the performance of machine learning models with classical queueing policies and examine how accurately they reproduce routing decisions under different operating conditions. Through this work, the student will gain experience in queueing simulation, data generation, machine learning, and performance evaluation. The project provides an accessible introduction to the growing intersection of operations research, artificial intelligence, and service system optimisation.

Project code:	SCI029
Project title:	**Talk data to me? Supporting English Language Learners in Statistics and Data Science**
Discipline:	Te Tari Tatauranga | Department of Statistics
Supervisor(s)	Liza Bolton
Contact details	liza.bolton@auckland.ac.nz
Skills Needed	• Strong statistical knowledge and vocabulary, likely from taking several statistics papers, ideally with at least one at stage three • Experience critically reviewing and synthesising information from

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

	academic literature • Interest in education research, equity and student support • Data analysis skills in R • Experience in, or interest in building skills in, qualitative analysis of transcripts a bonus • Experience with language learning a bonus

Project description

An ongoing project in our department is working on understanding the experiences and needs of students who are English Language Learners in statistics and data science.

This Summer Research project will contribute to additional literature review, supplementary quantitative and qualitative analyses, planning and development of resources grounded in research and the experiences of our students.

Project code:	SCI030
Project title:	**Data: The Final Frontier**
Discipline:	Te Tari Tatauranga | Department of Statistics
Supervisor(s)	Liza Bolton
Contact details	liza.bolton@auckland.ac.nz
Skills Needed	• Modern data technology skills with web scraping, APIs, data use ethics, as developed in a course like STATS 220 • Data documentation skills • R programming skills (especially loading and cleaning data) • R workflow skills (commenting code, R projects) • Basic experience with GitHub a bonus • An interest in space exploration a bonus

Project description

The National Aeronautics and Space Administration (NASA) provides a large range of data to the public, enabling a wide range of possible investigations and creative and informative outputs. For example, Hank Green made a website where you can see all the Artemis II photos arranged in a timeline, synced to the crew's timetable and position in space.

This Summer Research project will contribute to documenting the available data, access options (file types, APIs, web scraping) and usage guidelines, with a focus on considering image and text data and opportunities to make connections across data sets.

Project code:	SCI031
Project title:	**Using multivariate methods to characterise symptom patterns in ME/CFS and long COVID**
Discipline:	Department of Statistics
Supervisor(s)	Natalia Boven

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

	Anna Brooks
Contact details	natalia.boven@auckland.ac.nz
Skills Needed	• Familiarity with R programming • Familiarity with multivariate statistics is an advantage • An interest in health is an advantage

Project description

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a highly debilitating complex chronic illness that is commonly preceded by a viral infection. Long covid refers to ongoing symptoms following infection with SARS COV-2 (Covid-19), and a subset of people living with long covid meet criteria for ME/CFS. Both ME/CFS and long covid are associated with large numbers of symptoms across multiple body systems, likely reflecting diverse and interconnected underlying biological mechanisms.

Using survey data collected as part of a Liggins study on immune dysfunction in ME/CFS and long covid, this research project will use multivariate methods to identify and describe symptom clusters (i.e. symptoms which tend to co-occur) for these conditions. Patterns will be compared for people living with i) ME/CFS with onset unrelated to Covid-19 infection, ii) long covid who meet criteria for ME/CFS, and iii) long covid who do not meet criteria for ME/CFS. Symptom clusters will be related to known risk factors (e.g. previous symptomatic Epstein-Barr infection) and functional capacity.

Project code:	SCI032
Project title:	**Employment trajectories for people living with multiple sclerosis**
Discipline:	Statistics
Supervisor(s)	Natalia Boven
Contact details	natalia.boven@auckland.ac.nz
Skills Needed	• SAS or R programming • Regression techniques • Responsible and trust-worthy (as using IDI data)

Project description

Multiple sclerosis (MS) is a chronic autoimmune disease that impacts the central nervous system. MS commonly presents in early and mid-adulthood and can have a substantial impact on people's careers. This project will use linked administrative data accessed through the Integrated Data Infrastructure (IDI) to explore employment trajectories for people living with MS. Potential protective characteristics will be examined, such as access to high-efficacy disease-modifying treatments for MS.

As this research uses the IDI, Stats NZ will undertake reference checking for the scholar and provide confidentiality training. Analyses must be conducted within a secure data lab on campus (after approval is granted by Stats NZ). The cohort of people with service use suggestive of MS has already been constructed.

Project code:	SCI033

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

Project title:	**Questionable assessment practices: Scaling automated question models for large-scale statistics education**
Discipline:	Department of Statistics
Supervisor(s)	Anna Fergusson
Contact details	a.fergusson@auckland.ac.nz
Skills Needed	• Strong skills in R programming, including successful completion of STATS 220 and/or STATS380 • Experience with data cleaning and statistical analysis • Interest in educational data science or pedagogical design

Project description

Automated methods for generating questions for use in large-scale statistics courses are gaining popularity, particularly in light of ongoing developments in Generative AI. This project explores the design of "pedagogy-first" models for Automated Question Generation (AQG) used in our large introductory statistics course (~2,500 students each semester), with a focus on the existing question-generating model for mean and standard deviation estimation. Key tasks include: analysing the existing R code to reverse-engineer and articulate the pedagogical rules and design choices that define the model, conducting an analysis of historic student attempt data to determine if question distractors and options behave as intended, evaluating human-written versions of these questions against the articulated automation rules to identify consistency gaps, developing a framework to evaluate specific features of the generated dot plots (e.g., distribution shape, axis scales, etc.) to investigate how these visual properties may impact question difficulty. The insights from this graphical evaluation could lead to an exploration of AI approaches for analysing the data visualisations students reason with in statistical assessments.

Project code:	SCI034
Project title:	**Optimising response selection for comparative judgement of statistical writing**
Discipline:	Department of Statistics
Supervisor(s)	Anna Fergusson
Contact details	a.fergusson@auckland.ac.nz
Skills Needed	• Strong skills in R programming, including successful completion of STATS 220 and/or STATS380 • Experience with data cleaning and statistical analysis • Interest in unstructured text data and/or paired comparison models • Interest in educational data science or pedagogical design

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

Project description

Introductory statistics and data science students need to learn to produce short written explanations that are statistically sound and clearly communicated. In large-enrolment courses, comparative judgement offers a scalable way to engage students in evaluating writing, but its usefulness depends on which student responses are selected for comparison. This project will investigate how to choose a small but informative set of responses from a larger pool of anonymous submissions collected through the Quick, write! platform. Using historic paired comparison data, response selection strategies will be explored and evaluated to determine the stability and usefulness of comparative judgement outcomes. The project will involve analysis of large-scale unstructured text and interaction data, use of statistical models such as Bradley-Terry, and simulation or resampling methods to compare different subset selection approaches. There is also scope to explore simple rule-based or hybrid generative methods for classifying or pre-screening responses to support future scalable feedback systems. The project will contribute to the design of more effective real-time formative assessment of writing in large statistics classes.

Project code:	SCI035
Project title:	**Quantifying the discoverability and citation pathways of statistics education research**
Discipline:	Department of Statistics
Supervisor(s)	Anna Fergusson
Contact details	a.fergusson@auckland.ac.nz
Skills Needed	• Strong skills in R programming, including successful completion of STATS 220 and/or STATS380 • Experience with data cleaning and statistical analysis • Interest in network analysis, bibliometrics, large-scale text data (specifically references data) • Interest in educational data science

Project description

Statistics education research is published across journals and conference proceedings, but the discoverability of older conference papers is uneven because of historical differences in how proceedings were hosted and indexed. This project will investigate how publication infrastructure has shaped the visibility and citation pathways of statistics education scholarship, with a focus on IASE conference proceedings and SERJ articles. The student will help build a dataset of papers, metadata, indexing status, and citation links, and will use this to explore patterns of discoverability, citation impact, and scholarly connection over time. The project may include an audit of whether papers are indexed in Google Scholar, modelling of citation counts and indexing probability, and construction of a citation network to examine how conference proceedings and journal articles connect within the field. There is scope to compare statistics education with a related education field, such as computing education, where conference proceedings play a more central role. The project will contribute evidence to support improved publishing practices and will help clarify how dissemination systems affect the reach of educational research.

Project code:	SCI036

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

Project title:	**Adapting the saddlepoint method for discrete random variables**
Discipline:	Statistics
Supervisor(s)	Jesse Goodman
Contact details	jesse.goodman@auckland.ac.nz
Skills Needed	• Experience with R programming and simulation • Some knowledge of generating functions

Project description

The saddlepoint approximation is a systematic method for approximating an unknown density function in terms of a known moment generating function. It is useful when each individual in a large population contributes to a single random variable, and has often been used in statistical ecology.

The saddlepoint approximation works best for densities, when the underlying random variable is continuous. For discrete random variables, the traditional saddlepoint approximation works less well, and always fails at the boundary. This project will develop tools for automating saddlepoint techniques in terms probability generating functions, extending a code framework currently based around moment generating functions.

Experience with R programming and simulation would be a plus. The mathematical aspects of the saddlepoint approximation are not prerequisites, but mathematical applications could be explored as part of the project depending on the student. This project will also involve some of the optimizations needed for high-accuracy scientific programming and numerical calculation.

Project code:	SCI037
Project title:	**Summarising for Human to Human reading**
Discipline:	Statistics
Supervisor(s)	Thomas Lumley
Contact details	t.lumley@auckland.ac.nz
Skills Needed	• Programming skill in R (preferred) or Python

Project description

Public comment, eg to NZ legislation such as the recent Treaty Principles Bill or Regulatory Standards Bill, has been increasing in volume. Authorities are tempted to replace actual reading of public comments by LLM summaries, but there isn't good public trust of these summaries. This project looks at clustering the comments and taking random samples from each cluster as a way to summarise public input while still having real humans read comments submitted by real humans. You will apply some text clustering methods to public comments on US regulations (which are easily available) to seem what clustering methods seem to work well.

Project code:	SCI038
Project title:	**Modelling potential future inequity using population projections**
Discipline:	Statistics
Supervisor(s)	Andrew Sporle

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

Contact details	a.sporle@auckland.ac.nz
Skills Needed	• An interest in social or health statistics • Knowledge of basic algorithm design principles • An understanding of statistical variance and probability distributions • R coding skills • Data wrangling experience • Epidemiology skills preferred but not required

Project description

Social and health statistics are used to describe outcomes for whole populations as well as specific sub-populations such as age groups, birth cohorts or ethnic groups. These statistical outcome measure for sub-populations can be used to calculate absolute and relative inequity in social or health outcomes between population sub-groups.

This project will create code that uses official population and sub-population forecasts to forecast the future population and sub-population outcomes including inequity measures. The first part of the project involves using real population data for a single social or health outcome and recent StatsNZ population projections to create a template for doing the outcome forecasting and graphics output. The second part of the project involves creating ways for the template to be used for other outcomes.

Project code:	SCI039
Project title:	Data quality improvement tools for small official statistics agencies
Discipline:	Statistics
Supervisor(s)	Andrew Sporle
Contact details	a.sporle@auckland.ac.nz
Skills Needed	• An interest in social or health statistics • An interest in Pacific development or Pacific statistics • Knowledge of basic algorithm design principles • R coding skills • Data wrangling experience • Epidemiology skills preferred but not required

Project description

Official statistics agencies are required to produce statistics that meet predefined quality metrics for their local and international obligations. While large nations have the resources and skilled workforce to undertake this work, this is not the case for smaller nations working with much smaller budgets and limited access to skilled data technicians and statisticians. This project will create algorithms and resources for the production process of official statistics datasets using open-source tools. The work involves creating approaches for assessing quality issues including completeness, missingness, and data consistency that can be adapted to work in the production of a Pacific nation's official statistics. This work will involve working with a national statistical office of a Pacific nation as well as with the Pacific Community (www.spc.int).

Project code:	SCI040
Project title:	**Who are New Zealand's carers and how does caring impact their wellbeing?**

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

Discipline:	Te Huinga Hāpai Tāngata | COMPASS Research Centre (Social Science)
Supervisor(s)	Lisa Underwood
Contact details	l.underwood@auckland.ac.nz
Skills Needed	• Quantitative data analysis • Ability to transform raw data into graphs and summary tables • Using R to manage, transform and visualise data • An interest in producing public user friendly statistics outputs

Project description

Te Huinga Hāpai Tāngata COMPASS Research Centre uses quantitative methodologies to answer policy-relevant questions, to improve health, wellbeing, and equity in Aotearoa.

This project will contribute to work that is being carried out by COMPASS as part of the Mahi Aroha Carers' Strategy Action Plan work programme, which will inform Actions relating to Monitoring and The government data and evidence strategy for carers. The main focus will be synthesising quantitative evidence on the sociodemographic characteristics of carers and indicators of carers' wellbeing.

The student will join a multidisciplinary team to assist with the development of two outputs:

1) High level summary results of data relating to carers from the NZ Health Survey and the Youth, Heath & Wellbeing Study 2025. The student will be provided with aggregate data outputted from the Integrated Data Infrastructure (IDI) which they will transform and prepare for use in figures and tables.

2) An existing Shiny App that visualises the sociodemographic characteristics and geographic location of carers using Census 2023 data. The student will be provided with data and R code that will be used to adjust summary tables and figures within the current version of the Shiny App.

Project code:	SCI041
Project title:	**Information Theory for STATS 331 Examples**
Discipline:	Statistics
Supervisor(s)	Brendon Brewer
Contact details	bj.brewer@auckland.ac.nz
Skills Needed	• Experience with Bayesian inference (e.g. good grades in STATS 331) • Comfortable with programming in either R or Python

Faculty of Science
Summer Research Scholarships

2026/2027 Projects (Department of Statistics)

Project description

Whenever probability distributions are used to describe uncertainty about an unknown quantity, we can also describe the total amount of uncertainty present using the Shannon entropy, a central concept in information theory. In this project, the student will investigate the basic one-parameter examples from STATS 331 (Introduction to Bayesian Statistics) from the point of view of information theory. Instead of just calculating the posterior distribution given the specific data, we can calculate the entropy of the prior distribution, the typical entropy of the posterior distribution, and the mutual information, using direct numerical methods implemented in R or Python. The mutual information describes how much the data tells us about the parameters, in general, before knowing the specific data. We will investigate the dependence on sample size, and consider the implications for Bayesian experimental design.