UKCOTS24

University of Manchester, 13-14 June 2024

Book of Abstracts

UKCOTS24 Abstracts

1 CONTENTS

2 Presentation Abstracts

2.1 Thursday 13th June, 10.30 – 11.15, parallel sessions

2.1.1 Stats without maths

Teaching an accessible MSc level introduction to data science course without prerequisites

Author: Serveh Sharifi (University of Edinburgh)

This talk is about our experience in teaching an introductory data science course at Master's level without any assumed prerequisites beyond school-level maths. Titled "Insights Through Data", this course provides a basic background in data science skills and methodologies via

direct interaction with data through programming, statistics, and machine learning. The course is delivered in a fusion format for online and onsite students who have a wide range of backgrounds from art and social sciences to computer science. Python programming is taught in the course as a computational tool for learning and practising data science. In this talk, we will share our experience in teaching this course including experimenting with teaching programming from scratch (through collaborative pair-programming, leveraging tools like CodeRunner, Noteable and Jupyter notebooks and nbgrader), challenges of fusion teaching, and assuming no previous related background from students.

Stochastic processes for the masses: inspiration for an interactive introductory lecture suitable for open days, non-specialists and beyond

Author: Tom Honnor (UCL)

The spread of misinformation via social media, transmission of infectious diseases amongst people and animals, and the growth or shrinkage of national population sizes are all frequent topics in the news cycle. Each may be modelled at a base level by simple stochastic processes. This talk introduces a lecture plan used to introduce such stochastic processes to audiences with limited prior knowledge, as might be found at a typical open day or introductory statistics course for non-specialists, which has received positive responses from prior attendees, with potential to extend the scope for specialist students. Lecture attendees play an active role, using dice rolls to simulate the evolution of a stochastic process modelling the spread of misinformation online, and that interactivity will also be included in this conference talk. Attendee-generated simulations are complemented by the results of larger scale computer simulations to provide convincing evidence of process behaviour without the need for technical proofs, making the lecture accessible to individuals with only school-level mathematics education. Small changes to the stochastic process structure and parameters are shown to lead to the misinformation dying out or spreading out of control, with attendees guided towards an intuitive focus on the reproduction number, R. The breadth of applications of stochastic processes is further illustrated in the lecture through links to the spread of infectious diseases and changes in national population sizes. Lecture attendees leave the session with an insight into the breadth and importance of real-world applications of stochastic processes, the power of simple summary calculation results to understand the behaviour of complex processes, and the use of computational simulations to shed light on the behaviour of systems initially beyond analytical comprehension. Possible extensions for lecture audiences with more advanced prior knowledge could include formal derivation of results for the stochastic processes, exploration into how such processes can be programmed and information synthesised from computer simulation results, and consideration of the fact that the probabilities underlying application-focused stochastic processes are unknown in practice and need to be estimated statistically using observed data.

What no calculations?! A course focused on the application and communication of statistics

Authors: Catherine Palmer (Munster Technological University)

“I beseech you, in the bowels of Christ, think it possible you may be mistaken” Oliver Cromwell, 3rd August, 1650. Many third level statistics courses prioritize theory, technique, and analysis over the communication, dissemination, and impact of research findings in the real world. The

proposed talk describes student and lecturer experiences of delivering an alternative style of statistics course that focusses on assessing the trustworthiness of the statistics that are encountered in all domains and walks of life. 'Critical thinking with statistics' was part of a Masters in Investment Fund Administration offered jointly by South-East Technological University and Munster Technological University, Ireland. Tailored for part-time students who work in the financial services industry, the course was designed to equip students with the ability to think critically about statistics from a broader perspective. The aim was to encourage learners to consider the peer review publication system, and to help develop the skills to accurately evaluate the reliability, reasonableness, and truthfulness of statistical findings, claims and conclusions reported in the finance industry, the media, and the public sphere. This talk will outline the aims of the course, describe the content, delivery, and modes of assessment. The lecturers will reflect on their experiences and the challenges of teaching a statistics course with no theory and calculations. A summary of student self-reflection and feedback on their learning experiences will be presented along with considerations for improving future deliveries.

2.1.2 Medicine and allied health sciences (Burwalls)

New viruses are inevitable; pandemics are optional – lessons for and from statistics

Authors: James Nicholson (Statistical consultant), Jim Ridgway (University of Durham)

We explore ways in which statistics can be used to understand disease spread, and to support decision making by governments. "Past performance does not guarantee future results" - we hope. We discuss and show examples from the NSF funded COVID-Inspired Data Science Education through Epidemiology (CIDSEE) project. The Data Detectives Club is structured around a science-based adventure book for middle school youth by Pendred Noyce, titled Pandemics! NetLogo and CODAP are used to simulate the spread of different diseases (Spanish flu, Ebola, Covid, Polio etc.) to understand disease spread via illustrative stochastic processes and visualise how changing parameters in the simulation affects the outcomes, and develop an understanding of the scale of the variability of outcomes in such situations. In traditional statistics courses it is not common for students to consider the impact of policy decisions on societal outcomes – numerous opportunities exist in the project for exploring 'policy' of enforcing mitigation mechanisms - from compulsory vaccinations, to quarantine, for example, and other approaches like herd immunity. Decision making with limited information, and decision making when limited resources are available are explored in accessible, real world contexts, for example in evaluating contrasting approaches to monitoring disease spread (track and trace vs pooled testing). Throughout, the emphasis is on the relationships between evidence, modelling and theorising, and appropriate action. Statistics should be an essential element in all these aspects. We point to some 'big statistical ideas' that underpin the whole process of modelling, that can be illustrated vividly in the context of pandemics. We argue that statistics education should emphasise the application of statistics in practical situations, and that many curricula do not equip students to use their understandings of statistics outside the classroom. We offer a framework for curriculum analysis, and point to some rich teaching resources.

UKCOTS24 Abstracts

What and how are study design and analysis taught to undergraduates? A survey of four life-science disciplines across the UK

Authors: Crispin Jordan, Franziska McManus (University of Edinburgh)

Statistical analysis lies at the heart of inference for the life sciences. Yet, a reproducibility crisis across multiple disciplines raises the questions of whether, and how, teaching of statistics needs to change? To begin to investigate these questions, we approached every UK university to survey what, how much, and how statistics and experimental design are taught for four life sciences: Biology, Biomedical Science, Medicine and Psychology. Our survey aims to determine whether some topics are universally under-taught, and whether disciplines differ in ways that allow disciplines to learn from each other. We obtained 31, 19, 15 and 38 responses from Biology, Biomedical Science, Medicine and Psychology, respectively. The results reveal stark differences among disciplines in the amount of time spent teaching both study design and data analysis, where Psychology taught the most, and Biology tended to teach more than both Biomedical Sciences and Medicine. Psychology tended to provide more "stand-alone" courses that focus on study design or data analysis, while the remaining disciplines tended to integrate these topics into other non-specialized courses to a greater extent. The survey revealed broad diversity in whether statistics is taught from a mathematical vs. functional perspective (and little among-discipline difference in approach) - this finding highlights opportunities to identify approaches that minimise student 'maths-phobia'. With respect to teaching statistical analysis, all disciplines tended to teach a core set of similar statistical tests (e.g., t-tests, frequency (Chi-Square) tests, correlation, ANOVA); disciplines differed with respect to offering more advanced techniques (e.g., mixed effects models, PCA), where Psychology offered the most. Disciplines tended to teach similar materials with respect to Study Design. Strikingly, however, most disciplines tended to fail to teach topics with known links to low research reproducibility, including pseudo-replication (i.e, violating the assumption of independence), power analysis, and blinding. Further, all disciplines often did not teach about "Questionable Research Practices", which also likely lower research reproducibility, (e.g., HARKing, p-hacking, adding and excluding data). Similarly, all disciplines tended to omit topics related to Open Science (but Psychology tended to provide the most training). Our results highlight both topics that require more teaching across all disciplines, and opportunities for disciplines to learn from each other to improve training in study design and analysis. We suggest that courses switch emphasis from teaching "experimental design and analysis" to teaching "skills required to produce reliable research" and support undergraduates as future researchers.

2.1.3 Development and skills

Scholarship of teaching and learning in statistics education

Author: Rhys C Jones (Educational consultant)

This session explores the intersection of the Scholarship of Teaching and Learning (SoTL) and professional development for educators in the field of statistics. The presentation focuses on the impact of engaging in research and reflective practices related to statistics education on Continuing Professional Development (CPD) and career advancement. Drawing on established SoTL frameworks, the session addresses how educators can leverage their contributions to teaching and learning to enhance their pedagogical skills, contribute to the scholarship in their discipline, and support their promotion opportunities. The session will cover examples that specifically address the development of activities (such as small scale pedagogic based projects, areas of innovative practice identified through peer review, innovative course design)

that can lead to the development of meaningful and relevant research questions, study design, and methodologies for conducting meaningful SoTL research in statistics education. By offering insights into the symbiotic relationship between SoTL engagement and career progression, this session aims to inspire and guide educators in their pursuit of excellence in teaching statistics. Practical tips and advice will be given, drawn from my own experience of sitting on senior promotion panels as an associate dean for education in my previous roles, with additional significant experience of reviewing external applications for promotion all at levels (i.e. appointed as an external reviewer for professorial level appointments at other institutes across the globe etc), as well as chairing and sitting on numerous grant panels for funding for SoTL related pedagogical activities.

Maths is a language

Author: Theresa O'Brien (University of New South Wales)

Learning to read and write maths is a major barrier to students in statistics. In this talk I will address some linguistic structures in how we speak maths: notation and information density, grammatical processes, and a grammar of algebraic expressions. By making these structures explicit we can lower barriers to developing fluency in students, and improve stats communication to a more general audience by offering tools to break down and express complex models. This talk needs very little maths or grammar and is aimed at a maximally diverse audience.

Embedding sustainability into statistics education

Author: Riz Nawaz (University of Leeds)

There is an increasing desire for students in UK Higher Education to gain skills relevant to the sustainability agenda. Here, I provide an overview of how sustainability is being embedded into the teaching of statistics within a University Skills Centre. Since the provision is delivered centrally, there is an opportunity to engage with a diverse group of students to maximise reach.

2.2 Thursday 13th June, 11.45 – 13.00, parallel sessions

2.2.1 Teaching modelling and testing

The odds are it's wrong: correcting one of the most common mistakes in statistics

Authors: Laura Watt (Manchester Metropolitan University), David Voas (UCL)

Binary logistic regression is one of the most widely used methods in inferential statistics. We often want to predict a binary outcome: win/lose, pass/fail, mover/stayer, infected/not, and so on. Unfortunately the method uses odds, log odds and odds ratios, which are difficult to comprehend and interpret. Understanding of logistic regression tends to fall down in one of three ways:
1) Many students and researchers come to believe that an odds ratio translates directly into relative probabilities. The assertion that an exponentiated coefficient of 2.0 means that

increasing the variable by 1 makes the outcome twice as likely is easy to find in online teaching material, never mind student papers.
2) Alternatively, they learn that coefficients tell us whether the predictor variables make the outcome more or less likely, without knowing how to interpret changes in the odds. They will correctly write that "the odds for women are twice as high as the odds for men," but such a statement means little on its own.
3) They may be instructed in how to find predicted probabilities, but the calculations are laborious to perform by hand and further training is needed if the job is done using statistical software.
Our key aim is to highlight and correct the common mistake of confusing differences in odds with relative risks. Students can learn that coefficients tell us whether particular independent variables make the outcome more or less likely, but this approach is hardly very satisfying if they cannot interpret changes in the odds. Probabilities are much more helpful, and so we describe a simple method of estimating them for both binary and interval variables. We also try to show how odds, odds ratios, log odds and logistic regression can be introduced to students with limited mathematical backgrounds.

Do you see what I see? Challenges of interpreting model diagnostic plots

Author: Simon Taylor (University of Edinburgh)

Do you see what I see? Constructing data visualisations is an incredible useful tool to explore and assess structures within data that may be challenging to evaluate numerically. For example, creating a fitted against residual plot and a normal quantile-quantile plot is common practice to diagnose whether the underlying assumptions of a fitted linear regression model are being satisfied. Such images are designed such that it is visually obvious when modelling assumptions are being violated, but what about the cases when the assumptions are reasonable? Typical guidance such as 'there should not be any disenable pattern within the scatter of points' or 'the points should lie roughly along a line' are imprecise that often lead a highly subjective and over-interpreted decision. Training students to be robust interpreters of model diagnostic plots can be challenging. Not only do students need to be taught about how the visualisations are constructed and of their purpose, but also about how their structure could naturally vary under the statistical principal of repeated sampling. Supplementary features such as reference lines, smoothed curves and intervals can help aid interpretation, but introducing such features needs to be approach with care so that students are not tempted to use these as the sole decider for or against assumptions irrespective of other evidence in the visualisation to the contrary. Consequently, developing a skill in making a robust interpretation is best achieved through exposure, but are students exposed to enough variations to be sufficiently trained by the time they graduate? In this presentation I will present work conducted as part of an undergraduate final year project on investigating students' robustness in interpreting standard diagnostic plots for assessing the assumptions for a linear regression model from standard diagnostic plots, and the development of a Shiny Application to help introductory statistics students understand how sampling variability influence the structure of such images.

How should we teach "the world beyond $p < 0.05$" to non-statisticians?

Author: Peter Martin (UCL)

P-values and the concept of statistical significance are misunderstood by many students, researchers, and even by some teachers of statistics. The "p < 0.05" threshold and the associated cultures of research and publishing have been held partly responsible for a replication crisis in many scientific fields. One root of this malaise may be the way that statistical inference is often taught at university level to students in non-statistical disciplines. Our teaching team developed a new approach to teaching hypothesis tests that aims to integrate the principles of the American Statistical Association's "Guidelines for Assessment and Instruction in Statistics Education" (GAISE, Carver et al 2016) with Wasserstein et al's (2019) ATOM recommendations for good practice in statistical inference: "Accept uncertainty; be Thoughtful, Open, and Modest". We emphasize how statistical inference is used in real research in the students' own fields, demonstrate good practice, and show what can go wrong when p-values are misunderstood, misinterpreted, or intentionally gamed (p-hacking). Specifically, we suggest seven didactic principles that emphasize the real-world consequences of statistical inference:

1. Link Type 1 / Type 2 errors to direct replication
2. Link unobserved biases to conceptual replication
3. Explicitly challenge misconceptions about p-values and "significance"
4. Show how misuse of p-values leads to (publication) bias
5. Emphasize the role of statistical inference within the scientific process
6. Demonstrate the importance of estimation
7. Encourage thoughtfulness about hypotheses

This presentation will illustrate how we implemented these principles in teaching an intermediate MSc module in statistics for non-statisticians, using examples of empirical research from sociology, psychology, and epidemiology. We will also report on a student-led evaluation of the new teaching material, which involved in-depth discussion with two MSc students and focus groups with a further seven students. The evaluation highlighted the importance of six aspects of teaching delivery from the students' point of view: clarity, consistency, thoroughness, real-world relevance, liveliness and narrative, and integration of teaching across modules. The students' views about the importance of narrative – "a good story" that illustrates a technical point – supported our efforts to link the theory of hypothesis testing to case studies of its use in real-world research.

Language and images to promote conceptual understanding of p-values and of confidence intervals

Author: Hilary Watt (Imperial College London)

Objectives: Many students struggle to gain conceptual understanding of core statistical concepts. Teaching strategies will be presented that have earned excellent feedback. This includes that such approaches make statistics easy to understand and that they overcame statistics previously feeling counter-intuitive.
Strategies: Standard deviations: emphasis that larger things generally have larger standard deviations (much larger for elephants than for dogs than for mice). Then suggest that dogs are very variable, with major differences in size between breeds. Hence even if a group of cats and a group of dogs both have the same mean, we expect the group of dogs to have the higher standard deviation.
P-values: a graph that illustrates the continuous relationship between z-values and p-values. Z-values (for unpaired t-test) are the ratio of a difference in mean between groups to its precision of estimate. The precision of estimate (standard error) considers the difference in mean (between groups) in study participants to estimate the difference in mean in the population, assuming random sampling of participants from this population. A greater difference in

participants' group means, relative to its precision of estimate, implies a smaller p-value and stronger evidence that the population difference in group means is not zero.
Confidence intervals: illustrating confidence intervals for the mean on histograms of raw data, can help explain the association between variability and width of CIs. Larger random samples give greater precision of population estimates; the sample mean is an imprecise estimate of the population mean, with the confidence interval reflecting imprecision.
Transparent explanatory interpretations can be repeated throughout teaching to reinforce core concepts. For instance, $p<0.0001$ implies tiny/ negligible compatibility with random selection from a population where the association of interest does not exist. For CIs: The mean in participants is 4.5 mmol/L. When generalising beyond these participants, we are 95% confident that population mean is between 4.3 and 4.7 mmol/L.
Small-group then whole group discussion will be encouraged of further strategies and exercises that promote conceptual understanding.
Jokes might have some role in relaxing students, to help overcome stats anxiety, that can interfere with concentration. Jokes might potentially help students to break-through any misconceptions students may hold.

2.2.2 Medicine and allied health sciences (Burwalls)

Overview of statistics teaching within undergraduate programmes in UK and Ireland dental schools

Author: Sam Leary (University of Bristol)

The UK General Dental Council (GDC)'s intended learning outcomes for undergraduate dental students briefly mention critical appraisal, but not statistics. Hence there is likely to be wide variation in how statistics is taught across the Dental Schools, which has not yet been well documented. We have designed a questionnaire to capture the main features of each of the Bachelor of Dental Surgery (BDS or BChD) degrees in the 18 UK and Ireland Dental Schools in terms of statistics-related teaching. This includes statistics taught in stand-alone courses/modules, or as part of other courses/modules such as research methods, critical appraisal, and research projects. Some of the questions within this survey were developed from the overview of teaching of statistics within medicine and allied health sciences across UK universities (Overview of teaching of statistics within Medicine and allied health sciences across UK universities | The University of Edinburgh). Ethical approval for this questionnaire has been obtained from the University of Bristol. The questionnaire has been circulated to contacts that have already been established in each of the Dental Schools. Once data collection is complete we will be able to summarise the variation relating to areas such as: teaching methods used, statistical concepts covered, use of theory/formulae/statistical packages, student contact hours, integration in the curriculum, assessment methods used, staff involved, student feedback/perception and overall aim of the statistics-related teaching. A comprehensive review of undergraduate dental statistics provision will allow sharing of ideas between dental statistics teachers in different institutions. It should also enhance undergraduate dental education in terms of statistics-related teaching, providing evidence of the disparity of this teaching to the GDC. Future work will investigate dental graduate views on statistical learning needs for clinical practice, in order to identify and address disparities between what is taught in undergraduate programmes and what is needed in future careers.

Using peer assessment and self-reflection to teach statistical programming

UKCOTS24 Abstracts

Author: Laura Bonnett (University of Liverpool)

Computer programming is a vital skill for anyone involved in the analysis of data. One of the best ways to become familiar with software is to use it. However, ensuring that students devote sufficient independent study time to become proficient in at least one language before the end of their degree is challenging. Therefore, the University of Liverpool's MSc in Health Data Science is designing an assessment strategy for its coding module to support students to practice coding, without burdening teaching staff with excessive marking.
The module will introduce students to a number of different languages via interactive workshops. There are two planned assessments. The first will ask students to write and run code to answer a question using a supplied dataset. Submission of this will occur after the end of the semester. In our experience, this usually involves students attending workshops (usually without any thought of the module in-between sessions!) and then completing the assessment in a panic immediately before the deadline! This does not embed the required coding knowledge, and will not lead to skilled programmers. Therefore, to ensure continued learning of the software outside of directed time, we will also be asking students to submit work fortnightly.
The challenge of fortnightly submissions is the excessive demands on staff time to undertaken marking. Instead, we will be asking students to peer assess the submitted work. Students will work in small groups and take it in turns to provide feedback to each student in the group. This will ensure that students regularly practice their coding whilst also benefitting from the knowledge and support of other students in their group. An additional benefit is training and experience in providing feedback, a necessary skill to all future workforces. A taught session on writing meaningful and constructive feedback will be delivered early in the module to support students with this task.
The second assessment will be an assessment of an example of the feedback the student gave and an evidenced reflection on what they learned about coding, and about critical analysis from the process of being assessors. To ensure a focussed reflection, students will be introduced to the STARR (situation, task, action, result and reflection) method.
Both assessments will meet the University's requirement for authentic assessment and should provide students with an enjoyable as well as a useful first experience of coding for statistical analysis.

Interactive Demonstrations for Teaching Key Concepts in Science and Statistics to Dental Undergraduates

Author: Gavin Revie (University of Dundee)

This is a workshop where I will share some simple interactive demonstrations I have developed to illustrate 3 important concepts in research and statistics: the importance of avoiding uncorrected multiple comparisons, why students shouldn't just consider mean values but also dispersion when reaching conclusions, and finally the problem of "vote counting" and why meta-analysis is sometimes needed. These demonstrations serve as a useful wakeup-call that re-engages students when tackling a subject they often find dry. They can be conducted with minimal resources using only a coin, a dice and some note cards. Because of their unusualness they also leave a lasting impression which will stay with students long after they graduate. I deliver these as part of my research methods and statistics lectures at Dundee Dental School, but they would also be applicable to other fields.

2.2.3 Using software

Innovative and Interactive Statistics Teaching using Quarto

Author: Kristian Evans (Swansea University)

Quarto is an open-source system which is useful for creating scientific teaching resources, in particular statistical resources, in a variety of formats. Quarto is the next generation version of R Markdown, and it can produce content using R, Python, Julia and Observable JS. The proposal is to provide a talk and demonstration of using Quarto based on recent experiences of using the system for teaching Statistics to post-graduate Data Science students and to second-year undergraduate Mathematics students. A brief overview of Quarto and its capabilities would be provided, followed by examples of resources created by the system. Particular focus would be made on its advantages over more traditional approaches, which include executing common coding languages and interactivity. A live demonstration would also be provided together with an opportunity to ask questions.

Reproducibile statistics: jupyter notebooks, github and DOI

Author: Jools Kasmire (UK Data Service and University of Manchester)

In school, we are all told to show our work. But as researchers and academics, how often do we really show our work? Well, in the modern age, we can not only show our work but also make it VERY easy for others to understand it, inspect it, cite it, and further develop it! This workshop demonstrates how jupyter notebooks, github (or other version control repositories) and DOI can be used to make your statistics and data work fully reproducible and citeable!

Choose what you use: Teaching different statistical software at the same time

Author: Laura Vinton (University of York)

Within the Maths Skills Centre, the recurring question, "Which software is best?" is met with the standard advice: "Opt for the one you find comfortable and have the resources for." However, this guidance often leaves students at a crossroads, needing a clear direction. From undergraduate to postgraduate studies students may have been taught multiple softwares. Students have personal preferences; for example, some want to try coding, and some don't. External influences, such as supervisory recommendations, can cause the decision-making process to become complicated. A pertinent question arises from this: How important is it for students to master statistical tests in a single, specific software program?
To broaden the learning opportunities we offer, the University of York's Maths Skills Centre has trialled a new approach: conducting two-hour workshops where multiple statistical software platforms are taught concurrently. Initially the motivation to teach different softwares concurrently was to offer students more flexible attendance options as they would be able to attend any workshop regardless of their preferred software. Students from any degree program can sign up for our workshops, meaning each workshop is attended by a mix of undergraduates and postgraduates with varied backgrounds in statistics. Contrary to our initial expectations that participants would gravitate towards a single software, a notable proportion of students actively engaged with multiple platforms, appreciating this session's comprehensive format. This session will demonstrate our strategy, detailing how we have restructured our resources to facilitate the simultaneous teaching of SPSS, STATA, and RStudio within a singular workshop

framework. There will be the opportunity to discuss the advantages, and potential drawbacks of this multifaceted instructional approach, and how practically it can be implemented.

2.3 THURSDAY 13TH JUNE, 13.00 - 14.00, LUNCHTIME ACTIVITIES

2.3.1 Speed networking

Author: ROSE: Jenny Terry (University of Sussex), Paddy O'Connor (Queen's University Belfast)

Hosted by the Researchers of Statistics Education (RoSE) Network, this speed-networking session will facilitate quick introductions between delegates, giving participants the opportunity to connect with statistics education enthusiasts from different disciplines and backgrounds, across both higher education and professional training environments! This approach to networking takes some of the awkwardness out of it and could be especially beneficial for delegates that do not know others at the conference, helping to foster a sense of community.
Like most speed-dating events, participants will have about 5 minutes (depending on numbers on the day) to talk to a designated partner. After that time is up, participants in group A will remain seated, whilst participants in group B will move to the next table. Optional prompts will be provided to help keep the conversation flowing. Unlike speed-dating, you won't have to tell the organisers who you'd like to see again. Instead, quickly swap (virtual) business cards and/or catch up with your new stats ed BFF later in the conference.

2.3.2 Dentists statistics teachers' gathering

Author: Sam Leary (University of Bristol)

I have recently established a group of representatives from UK Dental Schools who are involved with teaching statistics/epidemiology/critical appraisal to undergraduate dental students, focusing on BDS degrees. Our overall aim is to enhance UK undergraduate dental education in terms of statistics/research methods teaching. So far we have had two online meetings (March and October 2023), and a smaller working group is currently developing a survey of the provision of statistics-related teaching for undergraduate dental students. Longer term, I hope to apply for us to become one of the Association for Dental Education in Europe Communities of Practice.
We would love to have a face to face meeting, and UKCOTS is the ideal opportunity. We can discuss the challenges we face and potential solutions, along with activities we would like to undertake, and how to take our group forward.
We would welcome anyone who has any interest in teaching statistics-related topics to undergraduate dental students to this session i.e. it would not be restricted to only those already in our Dental Statistics Teachers' Group.

UKCOTS24 Abstracts

2.4 Thursday 13th June, 14.00 - 14.30, Lecture

2.4.1 Alan Turing's Manchester

Author: Jonathan Swinton

No conference attendee will be startled at the name of the building in which this informal and site-specific talk takes place, though some may be surprised that through most of the twentieth century there was nothing on the University campus to mark the past presence of Alan Turing, a man now widely seen as a founder of computer science (and in a few quarters as an early Bayesian, and in even fewer as the creator of mathematical biology). I'll discuss Turing and his contemporaries in mathematics, including Phyllis Nicolson, Max Newman, Jack Good and Maurice Bartlett and in computing, such as Tom Kilburn and Audrey Bates. I'll illustrate the consequences on the built campus landscape and its hints for how changing attitudes to gender, class and sexuality have affected past and current storytelling about the electronic computing revolution. Jonathan Swinton trained in mathematical biology and for a time made his living as a statistician; his *Alan Turing's Manchester [manturing.net]* was published in 2018.

2.5 Thursday 13th June, 14.30 - 15.30, Plenary session

2.5.1 Research in teaching statistics in HE: getting started

Many statistics educators have not been trained in education research, therefore getting started can be both exciting and daunting. This panel session aims to demystify the process to guide aspiring researchers in statistics education. Conference attendees will suggest and vote on questions that will be posed to the panel of experienced statistics education researchers. Topics may include why do research in statistics education, what methodological approaches this research can take, as well as how to build collaborative networks, navigate ethical considerations, and publish findings.

Panel: Jamie Sergeant (chair), Jenny Terry (University of Sussex), Laura Le (University of Minnesota), Rhys C Jones (Educational consultant), Bruce Dunham (University of British Columbia)

2.6 Thursday 13th June, 16.00 - 17.15, Parallel sessions

2.6.1 Active learning/engagement

Evidence-based active learning

Author: Eilidh Jack (University of Glasgow)

Active learning puts students at the heart of the learning experience and is being increasingly utilised as a learning strategy throughout higher education. There is a significant body of evidence that points to the positive impact of active learning approaches on student

engagement, retention and attainment. The success of active learning generally centres on students taking responsibility for their own learning, often through engagement with materials and activities prior to in-person classes. This can present challenges as students may lack the discipline, motivation or time to consistently prepare for class. It is therefore crucial that as we move towards a more student-centred active learning approach, we are able to critically evaluate the success (or failure) of these new approaches. This talk will present results from a study which used learning analytics to evaluate a flipped learning approach implemented across multiple Statistics courses and discuss some learning and teaching interventions introduced as a result of this.

Evaluating Pedagogical Incentives in Undergraduate Computing: A Mixed Methods Approach Using Learning Analytics

Author: Laura Johnston (UCL)

In the evolving landscape of higher education post-COVID-19, this lightning talk will explore a comprehensive mixed-methods study that we conducted on a first-year undergraduate statistical computing module at University College London. The presentation will focus on the evaluation methods of newly introduced pedagogical incentives. The study meticulously assessed how these new incentives influence student engagement by combining learning analytics and qualitative data. Through this exploration, attendees will gain insights into the methodologies employed and the implications of the findings for enhancing pedagogical practices.
We will present a longitudinal map of resource interactions through Bayesian network analysis of Moodle clickstream logs from 204 students. The findings suggest that early interaction with course materials may indicate future engagement while revealing feedback loops incorporating the new pedagogical incentives, suggesting their potential to foster sustained engagement. Additionally, we will share insights from focus group discussions that offer an in-depth look into students' experiences with these pedagogical changes. The qualitative feedback provides a comprehensive understanding of the learners' perspectives on the incentives and practical implications.
The core of this talk aims to contribute an interpretable and actionable model of student engagement for educators, allowing them to evaluate and enhance their instructional approaches critically. By integrating objective, data-driven analysis with detailed student insights, this research offers a detailed view of engagement and the development of effective online pedagogical practices. While highlighting the importance of a mixed-methods approach, this presentation will underscore the complexity of capturing student behaviour in digital learning environments and how Bayesian networks can be a powerful tool for understanding and improving student engagement.

Group work: let’s work together to make it a success!

Author: Eilis Hannon (University of Exeter)

Group work is included as part of many higher education programmes as an attempt to get students to develop the communication and collaboration skills required in the workplace. It is also commonly used to encourage peer-to-peer learning and develop a sense of community within the course. However, it is often ineffective with some students frustrated by the experience due to lack of engagement from some group members. Furthermore, when used as

a form of assessment, the focus typically remains on evaluating the quality of the output (a presentation or report) and the knowledge/technical components of the programme rather than the soft skills are trying to foster through this activity. It is therefore questionable whether students really gain from it what we as educators intend.
The objective of this workshop is to develop a sharable and adaptable vignette for successful group work. We will revisit the primary objectives of a group work activity, develop mechanisms for success and reflect on how assessment criteria can align with these goals. Through this process, we will reflect on and evaluate our individual experiences to identify the barriers and opportunities. Together we will co-design a framework that will be sharable with the wider higher education community to enable us to incorporate group work into our academic programmes with confidence. This workshop is open to any educator from any programme, who has experience in using group work, has participated in group work or has an interest in using it in future teaching activities.

2.6.2 Medicine and allied health sciences (Burwalls)

Setting a problem without a right answer: Application exercises for Teams Based Learning (TBL) in Evidence Based Medicine

Author: Sarah Rhodes (University of Manchester)

The Medical School at the University of Manchester introduced Teams Based Learning (TBL) for first-year students from October 2023. A week of topic-based course content is brought together at the end of the week with a series of teams-based activities which ends in application exercises. Application exercises are designed to develop higher level thinking skills of application, analysis, evaluation and creativity on top of acquisition of knowledge. According to Roberson and Frachini (2014), a good application exercise asks students to make a decision after analysing a set of competing priorities and values in an authentic and complex scenario. To encourage student debate and reflective critical thinking, the exercise should ask teams to select the best option out of several plausible ones before simultaneously reporting and then defending their decision to the other teams (Parmelee, 2012). For medical education, application exercises commonly involve a patient case, and students are asked to choose, rank or discount options from a set of diagnoses or treatments (Parmelee and Michaelsen, 2010). Evidence Based Medicine (EBM) is an element of the medical curriculum where students learn about statistics, study design, critical appraisal and the implementation of research into practice. Recently, the task of incorporating EBM Application Exercises into the curriculum has begun. EBM lends itself well to the application exercise approach. Activities such as designing a research study, appraising the research of others or making a decision based on the results of a study will generally involve multiple competing points of view. The challenge is to design accessible and engaging application exercises that fit into the 10-15 minute TBL format and can be delivered simultaneously though an online learning environment to multiple rooms of medical students, facilitated by non-specialists. Experiences so far of developing application exercises to integrate EBM into the medical education curriculum will be presented. A short application exercise relating to the choice of control group in a randomised controlled trial will be set for the audience; they will work on it in small groups and defend their answers. The audience will then be asked to workshop their own application exercises on a topic of their choice to be entered in a shared document for post-conference use by others.

UKCOTS24 Abstracts

Teaching statistics to medics using a flipped classroom format

Author: Kasia Banas (University of Edinburgh)

At the University of Edinburgh, Research and Evidence Based Medicine (REBM) is delivered as a consistent stream of activity during the first two years of medical school training. In Year 2, REBM is taught using a flipped classroom approach, where students are given didactic materials to watch and read at home. The weekly in-person sessions are then delivered as practical labs and interactive tutorials, giving students an opportunity to practise and apply their learnings. The assessment for the module consists of 14 multiple choice questions included in a larger knowledge exam, and a coursework assessment comprising a data analysis report and a critical appraisal assignment.
A common criticism of the flipped classroom approach is that many students fail to engage with the assigned material before class, and thus come to the in-person sessions unprepared. In this study (n = 115), we used learning analytics to examine whether watching more videos or watching them before class was associated with better marks on the module assessments. We found that 90% of students watched the assigned videos before the weekly practical, and most students watched them just a day or two before class. We did not find any significant associations between our predictors of interest and the coursework mark. However, we did find significant associations with the exam marks.
When the number of MCQ exam questions answered correctly was used as the outcome, we saw a significant effect of the video watching time: $B = -0.007$, $p = .02$. This indicates that watching the videos one day later was associated with a decrease in the number of correctly answered questions, by a factor of 0.99 ($e^{-0.007} = 0.99$). We also found a significant association between the number of videos watched and the number of MCQ questions answered correctly ($B = 0.011$, $p < .001$), indicating that watching one additional video was associated with an increase in the number of correctly answered questions, by a factor of 1.01 ($e^{0.011} = 1.01$).
Overall, we found that watching more course videos and watching them earlier was associated with a better mark on the multiple choice exam. We did not find an effect on the coursework mark. This difference may be due to the fact that the MCQ exam questions drew more directly on the video content than the more open coursework assignments.

Feedback from MSc UCL Queen Square neuroscience students on flipped learning versus traditional learning methods in teaching medical statistics

Author: Saiful Islam (UCL)

Flipped learning is an educational strategy that encourages active student participation by moving traditional lectures outside the classroom. This approach enables students to delve into fundamental concepts independently and at their own pace, thereby fostering a deeper comprehension of the material. Additionally, it provides scheduling flexibility, allowing students to learn at times convenient to them, which is particularly advantageous for postgraduate students who juggle multiple commitments. By focusing on challenging areas and dedicating less time to familiar topics, students can customize their learning experience, enhancing its effectiveness. Complex concepts encountered during pre-lecture reading should be discussed in class sessions, contrasting with traditional learning settings, the educational process typically follows a structured format where students receive instruction during designated class time and are then assigned tasks or homework to complete outside of these sessions. The

emphasis in such settings often leans towards passive learning, where students primarily absorb information through lectures, textbooks, and supplementary instructional materials provided by the teacher. In this traditional model, students play a relatively passive role in the learning process, with the teacher serving as the primary source of knowledge and guidance. At the UCL Queen Square Institute of Neurology, which offers around eight distinct MSc programs attracting students from various academic backgrounds, designing a medical statistics module to accommodate this diversity is a significant challenge. During the academic year 2023/2024, a total of eight lectures were held for this module, with 200 students in attendance. The first five sessions employed flipped learning techniques, while the remaining three followed traditional teaching methods. Flipped learning sessions included collaborative lectures, group discussions, and Q&A sessions to encourage active student engagement and deeper understanding. We have collected feedback for all lectures separately and we analysed only four key questions for this presentation:
Rating of teaching quality (on a scale of 1 to 5, with 1 being poor and 5 being excellent), Clarity of learning objectives (Yes/No), Overall rating of the lecture (on a scale of 1 to 5, with 1 being poor and 5 being excellent), Additional comments on any aspect of the session (free text).
This presentation aims to illuminate the student experience with these two different teaching approaches.

2.6.3 Visualisation

Actionable visualisation principles and guidance for a foundational data science course

Author: David Sterratt (University of Edinburgh)

One learning outcome of the second year University of Edinburgh course "Informatics 2 - Foundations of Data Science" is that students can describe and apply good practices for visualising data. There are several sets of guidelines about what constitutes good visualisation practice in the literature and online, but each set focuses on different aspects of visualisation, there is no one-to-one mapping between the sets, and the level of the guidelines ranges from very general (e.g. "Show the data", Tufte, 1982) to very specific ("Avoid spaghetti charts", Schwabish, 2021). We could not find a single set of guidelines that was: (i) appropriate to the level of the course and the static visualisations we expected students to produce using Matplotlib and Seaborn; and (ii) actionable in the sense that students and markers could assess visualisations against the criteria. We therefore constructed our own set of five visualisation principles, each with a number of subsidiary guidelines. The principles and guidance is concise enough to be printed on an A4 sheet, which the students can use in group workshops to assess visualisations.
In this talk we will outline the principles and guidance, and how we use them in the course for instruction, formative feedback and assessment. We will focus on how we address many students' difficulty with creating legible text in plots, which means that many of them do not score well on the principle of making the data accessible. We will evaluate informally the efficacy of the guidelines.

Using StatSpace Visualisation Tools

Author: Bruce Dunham (University of British Columbia)

An obstacle for learners of statistics is that variation is difficult to convey in print and static images. Textbooks, for example, struggle to communicate the dynamic variation that is at the heart of the discipline. Interactive online visualisation tools can help learners to grasp difficult ideas such as the sampling distribution of the mean and confidence intervals, but only if the learners engage with the tools in the right ways. Here we present a suite of applets that are freely available in the resource repository StatSpace (https://statspace.elearning.ubc.ca), have accompanying activities, were beta-tested on students, and have been used from school level up to advanced courses for statistics undergraduates. The session will include a brief description of the development of the applets and discussions of both how to engage learners with such tools and proposals for further applets.

2.7 Friday 13th June, 9.30 – 10.30, workshops

2.7.1 Teaching statistics in an ever-changing technological landscape

Authors: Craig Alexander, Jennifer Gaskell, Vinny Davies (University of Glasgow). Elinor Jones (introduction)

The ever-developing landscape of modern technology has enabled impressive advances in the delivery of teaching in higher education. However, recent changes are so fast paced that educators are struggling to find an effective, pedagogically sound way of handling these advanced in their teaching and assessment. Technologies such as version control (e.g. GitHub), generative AI (e.g. ChatGPT), and modern machine learning techniques have quickly gained traction to great effect in recent years. With this increase in prominence also comes a demand for knowledge of these skills in the industrial and academic workplace, and the need to expose both staff and students to such practices.
In this session, we will discuss our experiences in sharing such knowledge with both student and staff learners, detailing some of the pedagogical challenges and considerations of integrating modern practices into the established statistics curriculum. We will discuss the challenges that must be considered when adapting assessment practices to technologies such as ChatGPT and lead an interactive discussion on differences in educator responses.
We will have three 15-minute talks followed by an open and interactive discussion session acting as a knowledge exchange of ideas, thoughts, and challenges.

• Experiences on creating and teaching a short course on version control – Craig Alexander, University of Glasgow
• Redesigning assessments in times of generative AI: a Python example – Jennifer Gaskell, University of Glasgow
• Training the Trainers: How do we educate staff in modern technologies? – Vinny Davies, University of Glasgow
Following the talks, we will host a 15-minute open discussion session.

2.7.2 Standard setting methods for a statistics exam paper

Authors: Vikki O'Neill, Dr Karen Cairns (Queen's University Belfast). Jamie Sergeant (introduction)

Following the onslaught of the COVID-19 pandemic, many assessments had to be rethought. The conventional in-person, invigilated, 2-3 hour written exam, consisting of a handful of “find the answer” or “prove this” questions, were deemed impossible. Thus, educators turned to new ideas, largely driven by the online options available to them at the time. However, as educators grappled with the complexities of designing these new assessments, the enemies of cheating and collusion raised their ugly heads higher.
Thus, the need to ensure fairness and consistency in exams became increasingly paramount, leading to the adoption of randomised questions in exams. Randomisation can be done in several ways; students taking the same questions in a different order; students taking a random subset of questions from a question ‘bank’; or students being randomised to different versions of the same exam [Noorbehbahani, 2022].
The notion of choice in exams has emerged as a topic of debate [Bramley & Crisp, 2019]. Using a case study approach, we examine an assessment administered to level 1 students enrolled in a "Introducing Probability" statistics module. The assessment, comprising six Very Short Answer questions, was conducted as a 60-minute, in-person, online invigilated exam. Each question had three versions (A, B, and C), resulting in 729 unique exam journeys, thereby introducing variability that necessitates robust standard setting methods.
While standard setting has long been established in medical assessments, its application in other fields warrants exploration and adaptation. This interactive workshop seeks to address the challenges and opportunities presented by the introduction of standard setting methods in statistical exam papers.
Participants will engage in collaborative discussions and hands-on activities to explore the suitability of different standard setting approaches for statistical exams. By critically evaluating the strengths and limitations of various methods, participants will contribute to the development of best practices in standard setting for statistical assessments.
Through this workshop, we aim to enhance participants' understanding of standard setting methodologies and their application in statistical exams. Furthermore, the collective expertise of participants will inform the development of guidelines for standard setting in statistical education, ensuring fairness and consistency across different exam versions.

2.7.3 Using playfulness to support student learning of quantitative data analysis

Authors: Julie Scott Jones, Carla Cordner, Simon Massey (Manchester Metropolitan University). Rhys C Jones (introduction)

This will be an interactive workshop suitable for any teacher of social statistics delivered by the Manchester Metropolitan University Q-Step Centre team. The Manchester Metropolitan University Centre pedagogic approach focuses on playfulness as a core element of its approach to teaching quantitative methods and statistics to social science students. Drawing on a decade of practice, supported by positive students’ outcomes, this workshop will outline how playfulness ‘works’ through addressing barriers to learning statistics for non-STEM students, building engagement, and supporting confidence with competence. The workshop will begin by briefly addressing some of the core barriers to learning that non-STEM students face when encountering statistics for the first time. This is not to emphasise a ‘deficit’ model approach but to provide a rationale for the use of a playful pedagogic approach. The emphasis will be on how a playful pedagogic approach can address these barriers, with its emphasis on ‘play’ as a means to apply learning, be creative and ‘have fun’. The pedagogic framework will be presented, alongside its underpinning theory and evidence to support its efficacy. The team will then, working with the audience, demonstrate how to implement this approach through three teaching activities: ‘balls to probability’, ‘what’s in the box’ and ‘telling stories’. Each

activity will illustrate a core concept(s) typically encountered in introductory statistics modules for social science students, including probability, sampling, parameters, population, measurement, representation. In addition, each activity will illustrate core skills required for introductory statistical analysis, including visualising data, framing data with theory, describing data. The adaptability of each activity to different learner levels and teaching spaces, will be emphasised. Each teaching activity will be interactive, and attendees will have opportunities to ask questions; teaching resources will be shared.

2.8 Friday 13th June, 11.00 – 12.15, parallel sessions

2.8.1 Assessment

An approach to formative feedback in statistics for a portfolio-based assessment

Author: Keith Harris (Sheffield Hallam University)

I would like to give a talk about the research I recently did about formative assessment and feedback as part of a summer project for Sheffield Hallam University's Academic Award for staff development. I investigated the pedagogic literature to find out what is considered best practice for formative assessment in other fields and looked into how formative assessment is approached in statistics instruction. This highlighted the need to create opportunities for practice and rehearsal before summative assessment (Sambell et al. (2013)) and to "close the loop" between students' thinking and the instructor's feedback (Lovett and Greenhouse (2000)). This research shaped my response to the requirement to move to a portfolio-based assessment for the third-year module I lead about Data Mining with Business Applications, as part of a drive to simplify assessment and improve formative feedback within my college. In my portfolio assessment, students obtain feedback on their initial attempt at some of the activities of the portfolio and are able to submit improved versions of these activities as part of their final portfolio of work. I will reflect on the benefits that these changes have made to my module, such as better technical understanding, enhanced inclusivity, and improved student performance and satisfaction, and identify potential challenges to replicating my approach, particularly for larger class sizes.

Research evaluating NLP tools designed to assist instructors with formative assessment for students in large-enrollment STEM education classes

Author: Matthew Beckman (Penn State University)

The project described here seeks to articulate the benefit of free-response tasks and timely formative assessment feedback and progress toward developing human-in-the-loop natural language processing (NLP) assisted feedback at scale. Research suggests "write-to-learn" tasks improve learning outcomes, yet constructed-response methods of formative assessment become unwieldy when class sizes grow large. If we are to pursue Statistics and Data Science Education across disciplines, we will surely encounter both opportunity and necessity to develop scalable solutions for pedagogical best practices. In a pilot study, several short-answer tasks completed by nearly 2000 introductory tertiary statistics students were evaluated by human raters and an NLP algorithm. The talk will conclude with recent developments building upon this pilot, as well as implications for teaching and future research.

UKCOTS24 Abstracts

The two Rs: teaching regression in psychology and neuroscience through reproducibility and replicability

Author: James Bartlett (University of Glasgow)

On our MSc Research Methods in Psychological Science programme, we have a statistics and research design course to teach students intermediate to advanced concepts such as regression, robust statistics, and Bayesian statistics. Traditionally, psychology students are taught inferential statistics using a toolbox approach of individual tests for specific scenarios. In this talk, I will outline our approach of focusing on a more flexible regression modelling approach to inferential statistics and share our experiences from running the course. I will also outline our iterative assessment method of students completing small homework assignments throughout the semester using open data from published psychology research. Our approach promotes key graduate skills in computational reproducibility, critical evaluation in critiquing the modelling approaches of published research, and communication through students reporting their findings.

Randomised summative e-assessment using R-exams

Author: Ilaria Bussoli (University of Bath)

In this lightning talk I will present highlights from a case study demonstrating the use of the "exams" R package to create a diverse range of randomised quiz questions in various learning management systems (e.g., Moodle, Canvas, Blackboard, etc.). I will focus on the usage of cloze type questions (e.g., questions that combine numerical, text and multi-response answers). The randomisation in this demonstration arises from the internal data, the sub-questions generated in the question and in the order in which the questions are presented.
I will present an example on hypothesis testing in a Moodle platform. This example will illustrate how to generate random data (e.g., normal data and known variance) and demonstrate how to randomise the test type required to answer the question (e.g., some questions will have a right tail test, others a left tail test, etc.) to allow each student to have a unique experience. I will also show how to code the correct answer(s) depending on the randomised hypothesis test occurred to the student.
In conclusion, I will briefly discuss how randomised questions can be integrated as a form of assessment in the UK system, and at the University of Bath in particular, and whether the "exam" package might have widespread benefit post-Covid19 in a HE UK setting.

2.8.2 Using software

Creating accessible and reproducible teaching materials with minimal scaffolding

Author: Clement Lee (Newcastle University)

"Where to put the solutions to the problem sheets" is an ubiquitous issue when creating teaching materials. While LaTeX provides solution modes, materials can only be generated in pdf format. On the other hand, there is increasing need on making materials available in other accessible formats, in particular html, giving tools like (R) Markdown more popularity. In this talk, we demonstrate how to generate both the question and solution versions of problem

sheets, in both pdf and html, with one single script in R Markdown or Quarto, and with minimal software scaffolding. Reproducibility is ensured as the code is embedded in the script, and executed in the process to produce the numerical and graphical results on the fly. If reproducibility is one of the teaching goals, this approach can also be used to preach what we practise. Lastly, while R is the language mainly supported by this approach, it can be extended to Python and possibly other languages.

AI Skills for Business: developing a competency framework for a resilient data science and AI profession

Author: Matthew Forshaw (Newcastle University)

Artificial Intelligence (AI) holds enormous potential for businesses, enhancing productivity and competitiveness. The UK's National AI Strategy sets out a vision to strengthen the UK's position as an AI and science superpower over the coming decade. A key part of this involves ensuring we build a diverse and talented workforce with the right AI skills for the future. As part of the National AI Strategy, the government committed to carrying out research into the skills needed to enable employees to use AI in a business setting, and identifying how national skills provision can meet those needs.
Led by Matt Forshaw (Newcastle University and The Alan Turing Institute), the Department for Science, Innovation and Technology's (DSIT) Office for AI has been working with InnovateUK BridgeAI, to develop research on the high-level competencies that businesses need their employees to engage with to enable AI adoption, including for traditionally non-technical roles. This work underpins a Government commitment to facilitate an increase in the number and diversity of employees across the UK workforce with access to relevant, high-quality AI training, ultimately addressing the skills barriers limiting AI adoption.
Dr Forshaw will reflect on his experiences leading the development of the UK's competency framework for Artificial Intelligence. He will describe his labour market intelligence work for Government, and advances in professionalisation of data science and AI occupations. This session will explore the roles of industry and academia in delivering the data and AI skills training required to equip the current and future workforce with skills required to confidently address current and future data and AI challenges.

JMP-start into teaching applied statistics

Author: Volker Kraft (JMP)

Today's curricula often require students to learn a range of analytics skills, which are critical for all practitioners who want to learn from data. With the right software, learning these skills can be hands-on and engaging, allowing students to explore and analyze realistic data without struggling with a clunky or tedious statistics tool. JMP is interactive and powerful point-and-click software for solving real-world problems. It is ideal for engaging, hands-on teaching of relevant data skills in various fields and is also used by scientists and engineers at leading companies across the globe.
While the fundamental skills addressed in this session include understanding variation and uncertainty, we will also look at applications like data modeling, designing experiments and quality management – all from a student's perspective.
This interactive session will demonstrate how JMP can help to engage students' curiosity and teach relevant data skills which are most in-demand in industry today. We'll guide you through a series of brief demonstrations, so that you can directly experience the difference JMP can

make for your course. Participants will receive a free license before the workshop, and the presenters will provide sample data and lead you through several hands-on examples in JMP. We will also discuss best practices and share resources to support integration into modern statistics courses.

2.8.3 Anxiety and support

Learning statistics: from anxiety to confidence

Author: Rachel Hilliam (Open University)

A second-year statistics module at the Open University has around 700 students studying it each year, from a range of qualification routes. When the module was originally written, it primarily served students on a Mathematics and/or Statistics degree – but in recent years, most students are Data Scientists, Economists or studying on the Open degree.
We are interested addressing the different needs of students by their qualification. Over recent years we have developed qualification-based tuition support where students are allocated to tutors on the basis of their qualification route. Students can attend a range of tutorials arranged by core topics, by qualification group, and on specific skills. We hoped that this approach might increase tutorial participation by building on the common ground of qualification route amongst students.
Most recently we have considered statistics anxiety and how this might affect our students. Existing statistics anxiety measures and tools focus on traditional face-to-face learning environments. This talk will outline our development of a self-reflection tool for recognising statistics anxiety in online and distance learning students. We aim to help our students to identify areas where they could increase their confidence in learning statistics, and to take action through linked workshops that are designed to guide students in improving their skills and knowledge in relevant areas.

Student Perceptions of Statistics Anxiety-Reducing Interventions – A Study That's Stat-tastic!

Author: Anna Riach (University of Leeds)

We have previously found that presenting students with content that raises awareness of statistics anxiety in a two-hour workshop, reduces statistics anxiety. However, we know there are students who are too anxious to attend such workshops (Marshall et al., 2022). Attempts to deliver the workshop as part of a timetabled class within a module have the advantage of reaching all students who are anxious about statistics. The disadvantage is that students who have positive feelings towards statistics may see little benefit in a whole hour of contact time being used for such a workshop. In this multi-institutional study, lecturers presented the statistics anxiety awareness content one slide per lecture or practical throughout a subject, in order to reach those that need it while not exasperating those that do not. Crucially, the effect of the "Stat-tastic" slides was compared to other features of teaching that the literature suggests lecturers can implement to reduce statistics anxiety, such as the use of humour, collecting own data, or being approachable. Some of our results surprised us and are useful for other statistics teachers deciding what aspects of their teaching to focus time and energy on to most effectively reduce students' anxiety. We also reflect on our experience of pre-registering

our study, the challenges of undertaking studies across multiple universities and the advantages this brings to the quality and applicability of pedagogical research.

Mathematics and statistics support beyond the timetabled hours: Results from a scoping review

Author: Emma Howard (Trinity College Dublin)

Many institutions in higher education provide students with additional mathematics and statistics support outside of their timetabled course hours. We define mathematics and statistics support (MSS) as any additional organised mathematical and/or statistical aid offered to higher education students, outside of their regular programme of teaching, by parties within the students' institution specifically assigned to give mathematical and/or statistical support. This interactive presentation will draw on a recent scoping review of MSS in higher education (Mullen et al., accepted), which found 148 publications reporting 136 studies that met this definition of MSS. The types of MSS provided include appointments, bridging courses, drop-in session, Maths Support Centres, online communication, paper resources, peer tutoring, tutorials, and workshops. The scoping review shows that the types of MSS favoured by the UK are Maths Support Centre, as well as drop-in sessions or booked one-to-one or small group appointments. The scoping review also shows that MSS is an effective approach to teaching students.The presentation will provide an overview of the results of the scoping review through displaying the results using the interactive R Shiny app Evidence Atlases (https://estech.shinyapps.io/eviatlas/; Haddaway et al., 2019). Evidence Atlas displays interactive geographical maps showing studies and their details. The app user can focus in on specific areas of the world, e.g., the UK, or colour the points representing studies according to a collected data category (e.g., MSS type). The format of the presentation is intended to be a guided discussion on the role of MSS in teaching statistics.

2.9 Friday 13th June, 12.15 – 13.15, lunch workshop

2.9.1 Connecting with CAUSE research

Authors: Laura Le (University of Minnesota), Matthew Beckmann (Penn State University), Michael Jiroutek (Campbell University)

Come join us for a conversation about research in statistics and data science education (SDSE) with the CAUSE Research group. We hope to generate both interest and ideas from the UKCOTS community, to welcome your involvement with CAUSE Research, and to learn how we can better serve statistics and data science education research(ers) together.
The CAUSE organization (causeweb.org) has been a vibrant and active community for approximately 20 years. CAUSE Research is a subsidiary entity of CAUSE that seeks to promote and support the advancement of statistics and data science education research and create an inclusive environment to engage researchers from a wide range of backgrounds, experiences, and disciplines. In part, we will discuss recent and ongoing initiatives, including CAUSE Research Satellite events in conjunction with US Conference on Teaching Statistics (USCOTS),

building interdisciplinary bridges (e.g., computing education), and a monthly CAUSE Research reading group.
For example, topics of conversation might include:
--How to connect with researchers from other disciplines (e.g., STEM education) and other SDSE organizations?
--What are the needs for supporting researchers in statistics and data science education?
--How can researchers be mindful of diversity, equity, and inclusion?

2.10 Friday 13th June, 13.15 – 14.00, parallel sessions

2.10.1 Active learning/engagement

Incentivising student engagement in statistics modules

Author: Chris Brignell (University of Nottingham)

One way to encourage student engagement is through incentivisation. Rewarding students with marks, even for relatively low-tariff assessments, can encourage good study habits in students who were previously disengaged with learning materials. However, using continuous assessment to drive engagement has inconclusive results on student learning. Therefore, this raises the question of whether alternative forms of incentivisation lead to better engagement and student learning. A more subtle form of incentivised engagement are active learning pedagogies. Game-based learning is a form of active learning where learning activities include recognised game formats such as quizzes or voting. The idea is that through fun and appealing learning activities, students are more likely to be actively engaged, and assimilate knowledge faster because there is an element of friendly competition with peers. In theory, gamification works by drawing student's attention to goals, nudging students in the right direction, giving students immediate feedback, rewarding good performance and breaking down learning into manageable tasks. However, not all studies demonstrate a positive impact on student learning.
In this study, students on statistics modules at an Australian and a UK university are incentivised through participation in games. At the Australian university students take part in quizzes and the leaderboard is displayed to the class. At the UK university students receive game points for attending class and completing homeworks. At the Australian university, incentivisation seemed to help academically weaker students to achieve higher marks in the final exam, while academically stronger students didn't participate in the game as much. At the UK university, incentivisation seemed to increase good student behaviour (attendance and homework completion) for around a third of students but this did not lead to better exam performance. At both universities, qualitative feedback from students was positive in terms of their enjoyment, providing extra motivation and assisting their learning. Overall, although game-based incentivisation enhanced engagement and better study habits this did not necessarily result in better learning. Incentivising learning and understanding directly, however, is harder to do because they are more difficult to measure. For students, it remains true that intrinsic motivation is essential when preparing for summative assessments.

Unlocking student engagement using collaborative keys

Authors: Michael Jiroutek (Campbell University), Laura Le (University of Minnesota), Steven Foti (University of Florida)

UKCOTS24 Abstracts

Collaborative Keys (CKs) are an active learning technique that provides students with the opportunity to work together to create an answer key for any assignment. A collaborative document (e.g., Google Docs) is used to provide continuous engagement in the learning process. Students jointly generate and refine answers while teachers provide real-time feedback. This approach is adaptable for any learning modality (e.g., in-person, hybrid, synchronous online, asynchronous online), student population, and learning assignment (e.g., activity, homework). It is also consistent with approaches to maximize student learning, such as alternative grading, low-stakes/low-stress assessments, and cooperative learning. Especially in online courses, CKs help foster a sense of community, with authentic interactions between students and between students and teachers. From our experience across multiple universities, with different student populations (undergraduate, graduate, and/or professional students in the health sciences), different class sizes (as small as 8 students and as large as 90 students), and different course formats (in-person, hybrid, synchronous online, asynchronous online), we advocate for the utility and adaptability of CKs by explaining how we each implement them. We will also provide examples of students' work and evaluation of their work, and resources and knowledge to inspire the implementation of CKs.

What data do you use?

Author: Jen Buckley (University of Manchester)

Data is important when it comes to teaching statistics. Data can help engage students and bring statistical concepts to life. But what data is used by lecturers and why do they use it? This lightening talk gives a quick run through of what the UK Data Service has learnt about how lecturers (mostly in the social sciences) use data in teaching. It will highlight how teachers establish 'what works' in a context of competing challenges and priorities. In exploring these issues, it will look to open the conversation to ask you what data do use and why?

2.10.2 Schools

Assessment in statistics at A level

Author: Neil Sheldon (Chair, Teaching Statistics Trust)

Assessment drives a great deal of teaching, particularly at school level. Good assessment can be valuable for developing understanding and technique. Poor assessment can make it difficult for teachers to deliver anything much beyond rote learning. This talk will look at recent A level examination questions in statistics, illustrating the good and the not-so-good. It will consider how assessment in statistics *should* develop over the next few years, particularly in the use of technology. This topic is of obvious interest to schoolteachers, but it is important for teachers in HE too in providing them with an understanding of how their students have been taught and assessed while at school.

UKCOTS24 Abstracts

2.10.3 Social Sciences

A 'step-change' in quantitative skills training.. what does that mean for mainstream social science programmes?

Author: Mark Brown (University of Manchester)

Responding to demands from employers and driven by initiatives like Q-Step there has been a push to increase the training of quantitative data skills in the undergraduate social science curriculum. For many Q-Step Centres developing a 'step change' in QM training included development of new specialist QM degree pathways that enabled suitably motivated students to combine a substantive Social Science with a more advanced training in stats and data skills. Whatever the merits of these pathways (and there are many), they impact a tiny fraction of the target population. While we need QM specialists many of the arguments that motivated Q-Step were about a need to raise the bar for the thousands graduating from mainstream social science programmes, for whom a basic statistical literacy is increasingly a requirement for entry into traditional graduate careers. Here the challenges are profoundly different. While QM specialist pathways open up curriculum space for an integrated and scaffolded approach to developing data skills, the reality on most mainstream programmes remains a packing of stats training into a single 20 credit core methods unit. Without extra curriculum space and given the very different starting points of students (in levels of interest, and anxiety over stats) the scope for 'raising the bar' can feel very limited on what is already for many students the most feared and least loved module of their programme.
So what's the way forward? The talk highlights two developments that are helping to reframe and revitalise the approach to the teaching of stats on mainstream programmes. One is the idea of 'quantitative embedding', whereby students encounter QM content and the application of statistics across the wider substantive curriculum. This immediately helps open up more curriculum space and time for programmes to deliver a more ambitious set of QM ILOs. More importantly it helps students to see this material as an integrated part of learning their discipline and becoming a social scientist. Both these outcomes help reframe the role of the methods module. The other development, which directly facilitates embedding, is the explosion of rich and accessible social data that can be used in teaching. When students can engage with real world datasets on the substantive themes that excite them, the learning of statistics and data skills can be energised and brought to life. The talk offers reflections on the efforts to develop a more embedded experience of learning statistics for social science students at the University of Manchester.

Developing foundational quantitative data skills: the UK Data Service training programme

Author: Vanessa Higgins (University of Manchester)

Foundational quantitative data skills are an essential element of statistical literacy for social scientists, but they are often overlooked in favour of more advanced methods. This session will explore the national programme of quantitative data skills training available from the UK Data Service. The programme provides a combination of synchronous online training events and asynchronous on-demand training materials targeted at a foundational level, with a focus on the skills needed to manage, find, access and use data assets such as Census data, government survey data and national longitudinal/cohort studies. These data, are valuable research resources that could be used more widely for social research and policymaking if quantitative data literacy skills were more widespread among social scientists. The

presentation will describe the purpose and reach of the training and will present recent developments including the results from a research project to explore the impact of the training on students who participate in it, a pilot asynchronous data skills training programme we have developed in partnership with three Doctoral Training Partnerships, and the development of a data skills framework for quantitative data skills training.

2.11 Friday 13th June, 14.00 – 15.00, plenary session

2.11.1 The future of UKCOTS

Panel: Elinor Jones (chair), Laura Le (University of Minnesota), Matthew Beckmann (Penn State University), Michael Jiroutek (Campbell University)

As we come to the end of the first UKCOTS, this final session will provide an opportunity to discuss its future. After a brief welcome and review of the conference by the Chair, our US-based contributors will provide a comparative perspective on the United States Conference on Teaching Statistics (USCOTS), which marked its 10th edition in 2023. There will then be an opportunity for delegates to discuss their views and overall impressions of UKCOTS, as well as future challenges and opportunities. Responses will be collated to help shape the strategy for subsequent UKCOTS meetings. We will close the conference with a look ahead to other upcoming statistics education events in the UK and further afield.

---

3 Poster Abstracts

3.1 Posters: all breaks with particular focus on Friday 13th June, 10.30 – 11.00

3.1.1 Fundamental statistical concepts and techniques in the biological and environmental sciences with jamovi

Author: Brad Duthie (University of Stirling)

Fundamental Statistical Concepts and Techniques in the Biological and Environmental Sciences: With jamovi is a new introductory textbook for learning statistics. It starts with the very basics and prioritises helping students develop a conceptual understanding of statistics and apply the most fundamental statistical tools. New concepts are introduced with examples designed to be familiar to the reader, serving as a useful starting point for exploring more abstract concepts. The book includes 26 chapters that introduce statistical concepts and nine chapters with exercises for working through practical skills in jamovi. Together, conceptual chapters and practical skills chapters could be used as one week of material in an introductory skills class. There are links to Rshiny applications throughout the text to help visualise statistical concepts. The book is freely available online (https://bradduthie.github.io/stats/), along with all datasets for completing practical exercises (https://osf.io/dxwyv). For teachers, test questions, data sets, and Rmarkdown feedback templates are available upon request.

3.1.2 What drives statistics anxiety in UK undergraduate psychology students

Author: Charlotte Connell (Cambridge University)

Statistics anxiety has long been identified and studied in the student population (Baloğlu, 2003; Onwuegbuzie & Wilson, 2003; Valle et al., 2021), with the Statistics Anxiety Rating Scale (STARS; Cruise et al., 1985) being widely used as a measure of statistics anxiety in previous research (Bourne, 2018; Siew et al., 2019). However, the STARS faces apparent difficulties including issues surrounding self-report and a potential bias towards finding negative attitudes towards statistics. The current study attempted to investigate the drivers of statistics anxiety by taking a participant-driven approach. The cross-sectional qualitative study used a virtual focus group with three undergraduate Psychology students who discussed their experiences of statistics anxiety during a Psychology course and the driving factors. Using a Thematic Analysis approach, (Braun & Clarke, 2006) we identified overarching themes of academia and internal barriers with reference to statistics anxiety. Our findings suggest that undergraduate Psychology students relate their statistics anxiety to: their lacking of knowledge surrounding statistics, a lack of accessible support, teaching and examination style, perceived worth of statistics, and their view of self. These findings provide a new insight as to why students continue to experience statistics anxiety beyond that offered by standard quantitative questionnaires. Going forward, teachers can respond to this by providing a supportive environment in which an emphasis on the worth of statistics and students' capabilities is placed.
In the poster, I hope to address the drivers of statistics anxiety among students and how these might be addressed by professors and by disciplines more generally.

3.1.3 Assessing inclusivity and accessibility in quantitative skills teaching at Masters level

Author: Christoper Mitchell (University of Liverpool)

Against a background of increasing financial pressure on universities, student recruitment has become an area of significant focus. Variable fees, increasing applicant numbers and a better informed pool of applicants have resulted in competition between institutions to attract students and between students for access to places at top universities. In this context, the School of Biosciences at the University of Liverpool recruits students from diverse backgrounds to both undergraduate and postgraduate programmes. This study examines the support and teaching on a Masters level statistics module. The module cohort is highly diverse, both in terms of mathematical and statistical ability, but also in terms of culture and undergraduate experience. As part of the School of Biosciences curriculum review, in line with the University of Liverpool 2031 strategic framework, the two Masters statistics modules currently offered will be rolled into one, advanced Masters statistics module. All Masters students will therefore now take a more advanced statistics module irrespective of background undergraduate degree. The aim of this study was to assess student attitudes to materials delivered currently, with a view to how these could be improved or supplemented in the advanced statistics course. We surveyed all students but were particularly keen to examine the opinions of the International students who had particularly struggled with the module content and assessment. We found that international students report a lack of confidence in quantitative skills topics relative to UK students, suggesting a difference in student experience on the module. Notably, a number of students, all from international backgrounds, reported no statistical or programming experience in their previous educational institutions. Assessment on the module was divided into 3 tasks; a poster, a data analysis report and a computer-based exam. During the module students reported a lack of confidence in completing the tasks, especially the international students who expressed uncertainty about elements such as formatting a scientific report and the mechanics of constructing a poster. Overall, the international students faced barriers to success on the module that were not experienced by home students, presenting issues for equality and inclusion in our quantitative skills provision. This will feed into the re-design of the advanced statistics module due to commence September 2024, ensuring inclusivity and appropriate support for all students. The development of inclusive, accessible support for quantitative skills is a persistent need in biosciences and beyond.

3.1.4 Strategies for promoting conceptual understanding of p-values, including of their continuous nature

Author: Hilary Watt (Imperial College London)

P-values: a graph that illustrates the continuous relationship between z-values and p-values. Z-values (for unpaired t-test) are the ratio of a difference in mean between groups to its precision of estimate. The precision of estimate (standard error) considers the difference in mean (between groups) in study participants to estimate the difference in mean in the population, assuming random sampling of participants from this population. A greater difference in participants' group means, relative to its precision of estimate, implies a smaller p-value and

stronger evidence that the population difference in group means is not zero.
Transparent explanatory interpretations can be repeated throughout teaching to reinforce core concepts. For instance, p<0.0001 implies tiny/ negligible compatibility with random selection from a population where the association of interest does not exist.
Exercises with results from randomised trials, with p-values and CIs, where the question is what to do next. An appropriate action might be more research, or to recommend the medication or to drop it.

3.1.5 Teaching data wrangling: key components required for valuable resources

Author: Hilary Watt (Imperial College London)

Teaching data wrangling: key components required for valuable resources
Objectives: preparing data for analysis is the most time-consuming part of many analysis projects. How can we make this process easier?
Key considerations:
Some students do not understand that how valuable is a tidy R script file. This may give them confidence to overwrite previous versions of their dataset. In practice, many students may keep multiple copies of their dataset. When challenged, many agreed that this does not work well for them. Hence it is valuable to teach them good file management.
It is valuable to reflect on what type of comments are most useful within R script files, to help students to edit them most efficiently.
Resources aligned around the workflow may help students to better prepare data for analysis. This may save time especially when they have no idea what search terms to use to find valuable content.
It is crucial that the results are correct. Hence it is valuable to include encouragement to check each step as we progress.
It may be valuable to make sure students are not spending lots of time cleaning variables that they never intend to use.
It is useful that students know whether it is appropriate to analyse huge datasets, or whether it there is excessive unnecessary power, with computer issues from excessive size. Or excessive controls for each case, which may hinder progress and make negligible contribution to power.
Summary: resources can include key features above.

3.1.6 Introducing Jamovi as an alternative to SPSS for teaching and consulting in Health Sciences

Author: Marijka Batterham (University of Wollongong)

IBM® SPSS® has been used for teaching and academic consulting in health sciences at the University of Wollongong, Australia. Initiated by lack of computer lab access during COVID-19, and an increased desire by students and staff to participate in online learning, it became necessary to find a cost-effective solution for provision of statistical software to enable students and academics to analyse data on their own devices. In addition, we wished to introduce our students to software they could use in their future careers, particularly if they worked in areas of health or policy where statistical software was not funded. We had previously conducted research on statistics anxiety in our health sciences students and found that higher perceived computational anxiety was negatively associated with subject marks in

biostatistics and an evaluation of our one-to-one statistical consulting service showed 50% of students and staff accessing the service struggled with statistical software. While we have successfully introduced R to many students and staff there has been a need for a more intuitive alternative. To address this, we introduced Jamovi (www.jamovi.org), a free open-source graphical user interface for R, into our teaching and consulting. Jamovi is easy to use, with a visual interface similar to SPSS. The output is dynamic, results and visualisations are updated if the data is altered. Menu options are straightforward, with specialized user developed modules available to be added. The package has been widely adopted by teaching and research staff and students. There are some limitations to Jamovi, firstly default plots cannot be edited, hindering the ability to have publication ready figures for differing journal or organisational requirements. Secondly, the menu options for computing and transforming variables are not simple for novice users. Other R based free statistical software is available, for example RK Ward (rkward.kde.org) and JASP (jasp-stats.org). JASP is similar in terms of ease of use to Jamovi and has more options for editing the output and could be considered by others looking for a free open-source R based menu driven statistical program.

3.1.7 Assessing statistical knowledge among health science faculty: implications for educator preparedness

Author: Matthew Hayat (Georgia State University)

This study investigates the statistical knowledge of faculty members within accredited health science schools in the US, focusing on the disciplines of dentistry, medicine, nursing, pharmacy, and public health. A stratified probability sampling approach was employed, targeting a diverse range of accredited US institutions. Invitations to participate in an online survey assessing fundamental statistical topics were extended to all faculty members within the selected schools. In total, 708 faculty members from 102 institutions contributed to the study, yielding an overall response rate of 6.5%. Key findings indicate that among faculty members engaged in teaching statistics, 17% had completed two or fewer courses in statistics or biostatistics. Analysis of the survey data revealed that, on average, faculty members who reported teaching statistics scored 84.7% on an eight-question multiple-choice assessment. Notably, 37.2% of these educators were unable to attain a grade higher than 'C'. In contrast, faculty members not engaged in teaching statistics scored an average of 62.1% on the assessment, with 77.7% failing to surpass a 'C' grade. These findings underscore the critical importance of statistical proficiency for researchers and scientists operating in the evidence-based Information Age. Moreover, they highlight significant gaps in the statistical training of health science educators, raising concerns about their capacity to instruct the next generation of health science professionals effectively. In response to these findings, this study offers recommendations for addressing the identified challenges, emphasizing the need for targeted interventions to enhance statistical training among health science faculty. By bolstering educators' statistical competencies, institutions can better equip them to prepare future healthcare professionals to navigate complex research landscapes with confidence and proficiency.

3.1.8 Attitudes towards Statistics with SATS-36 Survey: University of Edinburgh Case

Author: Ozan Evkaya (University of Edinburgh)

Students' attitudes towards statistics are important to understand for their learning process, since favourable attitudes lead to outstanding course achievement and positive consequences beyond the classroom. The Survey of Attitudes Toward Statistics (Schau et al., 1995) was used to understand the opinions of students enrolled on the introductory statistics course at Edinburgh University under six key factors: affect, cognitive competence, difficulty, value, interest and effort. Data was collected at the start and end of the course across two academic years (hybrid teaching in 2021-2022, and classical in-person classes in 2023-2024). Across different years, some key changes are underlined for further illustrations.
In this poster I shall present the results from our investigations into how student's attitudes change over the duration of the introductory statistics course and how their opinions may differ with regards to demographic and academic factors. Such quantitative findings can be beneficial to better understand the attitudes of our students and explore ways in which we may cultivate more positive opinions about statistics.

3.1.9 Statistical needs of linguistics students – perspective from undergraduate dissertations

Author: Patrycja Strycharczuk (University of Manchester)

Linguistics is a discipline that interfaces with psychology, social science and computer science, among others, all of which involve considerably different uses of statistics and quantitative methods. This can create challenges for incorporating quantitative training in the curriculum design, and the division of labour between general quantitative modules, and quantitative training embedded within discipline-specific modules. This contribution reflects on this issue from a position informed by a survey on the use of quantitative data in undergraduate student dissertations.
104 undergraduate dissertations were examined, representing five cohorts of University of Manchester Linguistics and English Language students graduating between 2017 and 2021. Of these, 77 (74%) were found to feature some quantitative component, whereas 60 (58%) featured a considerable quantitative component, defined as using quantitative data to support a crucial argument. 72 dissertations reported counts / frequency tables, and 72 included some form of data visualisation. A smaller proportion (32 dissertations; 30%) reported descriptive statistics, such as means/medians and standard deviations. Formal significance testing was used in 43 dissertations. Of these, 15 used a chi-square test, 10 used logistic regression, two used linear regression, and 16 used a test designed for comparing the central tendency in two or more groups (a t-test, Wilcoxon test, ANOVA, Mann-Whitney test). In terms of software, 61 dissertations (59%) used Excel, and 15 (14%) used R.
Overall, most dissertations engage with quantitative data in some way, which underscores the importance of integrating some quantitative training into the core curriculum. The particularly salient topics in this context include summarising and visualising categorical data, and the distinction between frequencies and percentages. Similarly, there was a relatively high use of methods based on specific research template, such as use of logistic regression to analyse sociolinguistic variation. In contrast, there was relatively little use of multiple linear regression in R, which is the foundation of quantitative training provided in the programme through a specialised elective unit, Quantitative Methods in Language Sciences. This suggests a gap between learning a statistical method and being able to apply it in a novel context, which could perhaps be addressed with reinforcement through guided class projects.

3.1.10 We teach what we learn: statistical education in UK undergraduate sport and exercise science-related degree programmes

Author: Tony Myers (Birmingham Newman University)

A large proportion of research findings may be false [1]. Underpowered studies [2], inappropriate research designs, and questionable research practices are at the heart of the problem — this includes sport research [3]. Many researchers misunderstand fundamental aspects of statistics [4] and it is important to consider how this affects the teaching of undergraduates who will enter the academic or industry workforce. To examine this, we surveyed 94 academics from 60 UK institutions on aspects related to the teaching of research methods and statistics on undergraduate sport and exercise science courses. Topics '*covered in depth'*, on average, included null-hypothesis significance testing, *p*-values, statistical significance, and assumption checks. This was at the expense of *'not covered'* topics such as the Neyman-Pearson approach, the Fisherian approach, Bayesian approaches, open science, pre-registration, and Registered Reports. Students were mostly assessed via written assessments (54%), although individual research projects (29%), multiple choice tests (29%), and computer-based assessments (27%) were also common. Most courses used SPSS (76%) and Microsoft Excel (64%), although open-source software such as JASP (20%) and JAMOVI (10%) were also used. Online videos and tutorials (73%) were referred to more than textbooks (68%) or journal articles (64%). The median response for modules providing students with the practical ability to conduct data analyses was 'strongly agreed', with 'agreed' the median response for providing effective preparation for dissertations and conceptual understanding, and 'slightly agreed' for students being properly prepared for data analysis in both sporting and industry settings. Content analysis of open responses indicated that participants judged the effectiveness of their teaching based on aspects relating to feedback (students, other staff), assessment (formative, summative), industry (preparedness for work), and future study (preparedness for dissertations/postgrad). The usefulness of student module evaluations was debated. There were also contrasting views on the relative effectiveness of students passing tests compared to developing conceptual understanding. There were some collaborative projects with industry, but responses generally indicated that students were not prepared to work in industry, with participants not even sure what that would require. Our study suggests that academics in our discipline mostly take a 'well-worn path' to teaching statistics. Although some alternative approaches are being used, staff are focusing on the 'doing' of statistics — conducting statistical tests — instead of developing conceptual understanding. Our study also suggests that particular emphasis should be given to the constructive alignment between academia and industry.

3.1.11 ASA DataFest 2024 at Edinburgh

Author: Vanda Inacio (University of Edinburgh)

The American Statistical Association (ASA) DataFest is a celebration of data in which teams of undergraduate students work on a large, complex, and surprise dataset over a weekend to find and share meaning and insights. The dataset and the work are likely beyond the scope of what students see in their courses. Founded at UCLA in 2011, ASA DataFest has experienced rapid growth over the years. It is now hosted by many of the USA's most prestigious colleges and universities, as well as several renowned foreign institutions. This friendly competition provides students with invaluable real-world expression, an opportunity to showcase their skills and

explore a data scientist's job, and a platform to network with professionals and peers. At the local run of this event in Edinburgh in March 2024, all UG students from the University of Edinburgh and Heriot-Watt University, with an interest in data, were invited to join. During the competition, academic staff, PhD students, and data scientists from industry guided students in their work. The event ended with brief presentations of the teams' work and assessing the students' ability to communicate results clearly and effectively. Teams were awarded in various categories, such as "Best Insights", "Best Visualisation", and "Best Use of Outside Data". This poster summarises our experience in running this event and how it can benefit students learning in data science.

3.1.12 Establishing a real-world patient dataset to teach data skills and statistics to medical students at the University of Leeds

Author: Wendy Harrison (University of Leeds)

Background: A self-selected student project has been introduced for second year medical students in 2023/24, using anonymised data from Leeds Teaching Hospitals NHS Trust (LTHT) Patient Pathway Management system (PPM+) which captures routinely collected patient data. As part of the 'ENQUIRE' curriculum review, this dataset will support students' learning in essential statistics, including how to analyse, interpret, synthesise and present data which can aid clinical reasoning.

Dataset: The fully anonymised and de-identified dataset includes variables relating to patient factors, sociodemographic factors, disease-specific information and health outcomes. Data are strictly confidential within the terms of a Data Protection Contract between LTHT and the University of Leeds. As such, students are required to access and analyse the data through a secure University network, and to complete appropriate Information Governance training.

Data skills and statistics: PPM+ data are used to explore issues around data collection, data quality and bias, to describe how 'big data' is used for research, audit and evaluation, and to practice statistical skills. In Year 2, the focus is on linear models, use of Stata, hypothesis testing and clinical interpretation. Lectures introduce key concepts and tutor-led small group sessions allow students to practice techniques using PPM+ data. Students are assessed through an oral presentation and a final report.

Evaluation: The previous student cohort were asked to respond to three key evaluation questions: 59% agreed that this module helped them to develop critical thinking skills and understanding for use in medical practice; 37% agreed that skills developed through the project will help them to make clinical decisions; and 76% understood how the project data can be used to answer research, audit and evaluation questions. The end-of-year evaluation for 2023/24 is currently in progress and results are expected at the end of May 2024.

As an interim evaluation, the current student cohort were invited to complete a survey in November 2023: 25% felt confident about analysing the PPM+ data for their research question; 41% felt prepared for their oral presentation; and 22% felt equipped to produce their final report. Free text responses indicated that the majority of students were grateful for the opportunity to access real-world patient data, but that delays in receiving the full dataset led to a lack of confidence around how to conduct and interpret their analyses.

Results from these evaluations will inform teaching delivery in 2024/25.

3.1.13 Stealing from the humanities

Author: Zak Varty (Imperial College London)

This poster will discuss how the teaching of statistics can be improved by adapting, borrowing or downright stealing teaching methods from the arts and humanities. We will focus on how student engagement was impacted by using a structured reading summary called a rhetorical precis. Through this example, I hope to encourage you to experiment with borrowing and sharing pedagogical techniques across the sciences and humanities. This might appeal to anyone who supports learners during self-paced study or wants to promote student engagement with academic journal articles.