CASE STUDY

Applied Data Analysis: A Problem-based Learning Approach

Ann Smith, School of Computing and Engineering, University of Huddersfield, UK.
Email: a.smith@hud.ac.uk

Abstract

This paper examines the transition of a conventional multivariate statistics module to a problem-based learning module, first implemented in 2021. The primary objective was to enhance students' problem-solving skills, bridging the gap between mathematical concepts and real-world applications. The approach was implemented to instil a deeper understanding of real-world data analysis, emphasising the interpretation of domain specific problems in mathematical terms and the production of reports for industrial stakeholders.

Findings indicate that the integration of problem-based methods not only improved students' comprehension of statistical techniques but also fostered a more profound appreciation for their practical utility in diverse professional contexts. The problem-solving cycle, a central component of the approach, guided students in critically analysing complex challenges and formulating data driven solutions. Furthermore, this study emphasises the potential for replicating the industrial study group experience within an undergraduate teaching environment.

Adopting a problem-based learning approach in the teaching of data analysis empowers students to apply their analytical skills effectively to real-world scenarios, strengthening their capacity to communicate insights and solutions to industrial stakeholders. The study underscores the value of aligning educational practices with the demands of data-driven industries, providing students with a competitive advantage in future research and the job market. The study is descriptive and reflective in nature.

Keywords: problem-based learning, knowledge exchange, Study Groups for Industry, multivariate statistics, applied data analysis.

1. Introduction

For continued industrial advancement and economic growth, the UK requires an informed pipeline of graduates in mathematics related disciplines to feed roles in research and employment. The Bond report (Bond, 2018) outlined several recommendations to achieve this aim including incentives for academic engagement with small to medium enterprises (SMEs) and resources for industrial engagement via study groups for example. Established centres for doctoral training (CDTs) for postgraduate researchers and at least one centre of excellence to foster direct engagement between mathematicians and industrial partners.

Critical real-world problems can all be solved using mathematical techniques, it is therefore imperative that university education is aligned to inspire the new generation of enthusiastic and creative researchers desperately needed. Current learning tends to be narrowly focused on theoretical aspects and often overly sanitised, historical data sets. Little or no consideration is given to the wider issues presented on application and execution in non-mathematical domains.

Knowledge exchange (KE) describes a two (or multi) way exchange of information and skills between academia and practitioners be they industrial partners, policymakers, or service providers. Critical skills include understanding the stakeholder requirements, formulating a mathematical problem and communicating findings in domain appropriate language.

2. Undergraduate mathematics

Typically, the focus of undergraduate teaching in the mathematical sciences is on the more theoretical aspects of topics, theorems and proofs. Whilst this supports greater understanding of processes in general, model assumptions and error estimation for example, it does not incorporate all required elements of mathematical applications in industry.

On validation of the BSc Mathematics degree at the University of Huddersfield in 2018, the course was designed with four distinct streams of learning. Namely, the calculus, statistics, technical competencies (including programming and linear algebra) and the professional streams. The professional stream includes a year one problem solving module where students study real industry problems and present their findings in a report to the stakeholder, a group project in year two and final year individual project. Initially the final year module Applied Data Analysis (ADA), within the statistics stream, was designed as a traditional multivariate statistics module, the theoretical partner to the 'Big Data Analysis' module which runs in parallel. However, on anticipating the first final year cohort (2021/22) it was abundantly clear the module was outdated. During the four years from validation there had been not just the Covid 19 pandemic but an exponential rise in the need for appropriately skilled data analysts and data scientists. Throughout this period, I became increasingly involved in KE in particular study groups for industry, both online hosted by the award-winning Virtual Knowledge Exchange in Mathematical Sciences (VKEMS) team, funded through the Rapid Assistance in modelling the Pandemic (RAMP) continuity network and in person, e.g., European Study Groups for Industry (ESGI). Study groups require the facilitator at least to engage with the entire problem solving cycle: formulating the problem mathematically to reporting to industrial stakeholders.

Traditionally we teach only a fraction of the necessary skills, most crucially students are generally not involved with formulating the research problem and communicating findings in non-technical language.

As the primary aim is to equip undergraduate students with the necessary skills to solve real-world problems it was increasingly apparent a change of teaching approach was necessary. None of the key problem-solving skills are trivial and require nurturing alongside the more commonly taught elements such as using computing packages. It was with this in mind, and many study group experiences, the decision was made to restructure the module and hence the assessment regime to reflect realistic procedures and meet industry needs.

Hence the decision to switch to a problem-based learning format and replicate study groups and the problem solving cycle (figure 1) as illustrated in Spiegelhalter (2019).

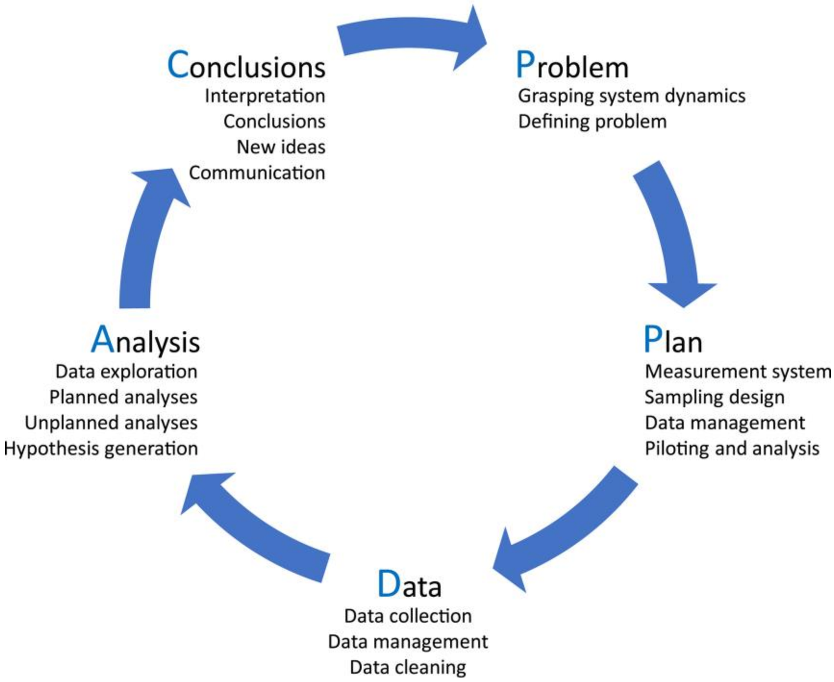

Figure 1. The Problem Solving Cycle PPDAC.

3. Problem based learning

Problem based learning (PBL) is an inherently student-centred pedagogy whereby students learn through the experience of solving open ended problems, pioneered by Barrows (1996) and extensively researched since, notably by De Graf et.al. (2003) and Wood (2003). The final year ADA module is ideally suited to a PBL approach. Additionally, PBL offers an opportunity to replicate study groups within the undergraduate taught programme so enhancing fundamental research skills.

Advantages and disadvantages of PBL are well documented (Newman, 2005; Yew and Goh, 2016) and include deeper understanding of techniques and development of critical thinking skills through active learning and application to real-world problems in varied domains. Working collaboratively fosters communication another key skill nurtured in study groups and introduced to the module teaching. Drawbacks of PBL relate to difficulties in problem choice and design to enable full coverage of the syllabus. These are overcome by the module leader’s frequent involvement in study groups for industry and wider research activity. Although suitably redacted, sensitive information omitted, and problems adapted to complement level of study.

4. Study Groups for Industry

Study groups, initiated by Oxford in 1968, are a long-established means of KE. Groups of academics and sector experts work intensively on stakeholder problems for short periods, typically one week maximum. Following which problem findings are presented as a technical report to the industry partner and, usually, submitted to a repository such as *Mathematics in Industry Reports* (MIIR) (n.d.).

Study group problems vary considerably in both domain knowledge required and mathematical area. During the period between enrolment of our first mathematics cohort and their progression to final year, I engaged with many sectors and industrial experts including local authorities, NHS providers, the hospitality industry, SMEs, and government departments. My experiences include an equally broad range of mathematical applications, some evidently mathematical such as ‘Estimating the risk of satellite collisions in densely populated orbit shells’ (Transfinite Systems ltd. ESGI # 162; July 2020) and others less so. For example, the City of Edinburgh Council problem ‘Homelessness prevention’ in January 2022 (V-KEMS Study Group, 2022a) and more recently modelling tourism flows in the city of Edinburgh for ‘Traveltech for Scotland’ (ESGI # 171, July 2023).

Formulating a mathematical problem in a social sciences context presents different challenges to those of applications in physics or engineering domains. Although in any application time spent understanding the relevant processes and fundamental definitions is crucial to understanding the problem and identifying potential data sources. Homelessness for example has a different specification for the local authority than, say the Shelter (2024) charity definition. Further, a large proportion of the homeless are not registered so suitable proxy variables need to be identified if possible. These issues are not isolated but are typical of real-world practice and should be incorporated within any applied data analysis course to better prepare graduates. Also noteworthy is the need for mutual respect between experts from all sectors, and communication in its broadest sense. Without good listening skills there is a danger the actual problem request is swamped by prior assumptions (Brewer, 1999; Brewer and Lövgren, 1999).

The VSG ‘Public Perceptions of Science’ (V-KEMS Study Group, 2022) aimed to bring mathematical scientists and other disciplines together to solve end user defined challenges with the aim of addressing the issues associated with the public perception of science. Problems were presented by the OSR (Office for Statistics Regulation), *The Times* newspaper’s scientific correspondent and the Winton Centre for Risk and Evidence Communication, Cambridge the need for improved communication skills for scientists at all levels was highlighted. The ability to disseminate information appropriate to the audience, in non-technical language, must be addressed. Additionally, a need for media training was identified as essential. Themes further expanded in January 2023 workshop ‘Communicating Mathematics’ (INI, University of Cambridge, January 2023). Furthermore, effective communication is critical to success of projects during the research phases, exercising intellectual humility and avoiding academic arrogance is key to initiating truly collaborative research with experts from other fields and domains (Krumrei-Mancuso et al., 2020; Porter and Schumann, 2018). Abandoning prior assumptions and prejudices is crucial to blend interdisciplinary expertise with maximum potential. Intellectual humility’s impact on learning is naturally also apparent, recipients more receptive to constructive criticism and benefit accordingly (Wong and Wong, 2021) with evidence of high grade averages.

5. Compelling evidence for educational updates

Misconceptions during COVID 19 pandemic, especially the difficulties in communicating uncertainty in predictive models, in conjunction with findings from the VSGs ‘Public Perceptions of Science’ and ‘Homelessness prevention’ motivated a significant change in approach to the ADA module. As detailed in the Bond report *The era of mathematics* (Bond, 2018) there are a rising number of challenges that require mathematics, hence a need to address the shortfall in appropriately qualified individuals. Further, education should meet industry requirements in equipping graduates with the necessary skills to tackle read-world problems.

Moreover, with the rising culture of fake news and conspiracy theories, deliberate and accidental miscommunication and misinterpretation of scientific findings, there is a need for the scientist to effectively articulate consequences and communicate actions to industrial practitioners and other non-subject-specific interested parties.

Incorporating discussions around topics currently reported in the news and on social media adds even greater relevance to study material. The current student population is more diverse than ever before and presents a rich blend of views and perspectives. Additionally, creating a relationship rich community (Felten and Lambert, 2020) fostering belonging, well-being, motivation and success. To this end, and to highlight the growing need to fact check reported statistics in the news, weekly discussions based on the More or Less podcast (BBC, 2024) were introduced and proved extremely popular. Further iterations also included comparisons of communications styles and formats and critical reviews. These were usually issued in advance of sessions and discussion as a class and in groups facilitated. More specific KE was also added as a workshop.

6. Practicalities

Although the module remains a course in multivariate statistics fundamentally, it has been repackaged to include the essential research and communications skills necessary for practical application of techniques taught. The reading list was extended accordingly to include report writing skills (Murray, 2017) and (Freiberger and Thomas, 2023). David Spiegelhalter's *The Art of Statistics* (Speigelalter, 2019) was also highly recommended. With just two two-hour sessions a week the students are expected to prepare material in advance in a semi-flipped style, thus allowing more time for debates and optimising the professional decision processes. I didn't throw the toy data sets away, they were used firstly with reduced dimensionality for pen and paper calculations, then with full dimensionality for MATLAB demonstrations and practical exercises using linear algebra (LA) style model constructions (Chartier, 2015). Subsequently, resources were reviewed, and practical applications were executed on chosen real-life data sets for example from the *Government Open Data* (n.d.) resources.

7. Summative Assessment Update

Clearly, with the change in delivery and adoption of a problem-based learning approach, there was a need to also update the assessment structure. Initially an equal split between a single case study assignment and examination, updating to a proposed series of reports required application to the School Validation Assessment Panel, with pedagogical justification for the changes. A necessary inconvenience to gain maximum benefit from the new module teaching format. Once approved, assessment was updated to a series of three written reports roughly equally distributed across the teaching period. Turnaround is tight to allow sufficient feedback discussion and implementation of related improvements prior to the next submission date. A more extended period of reflection would be ideal and currently consideration is being given to linking the first and second coursework with a slightly earlier submission date for the first to assess understanding, resource choice and problem formulation prior to further analysis and full report at the second submission point as illustrated in figure 2. Submission dates are equally spaced across the twelve-week period, coursework 1 in week 4, coursework 2 in week 8 and coursework 3 in week 12, and contribute 30%, 30% and 40% respectively.

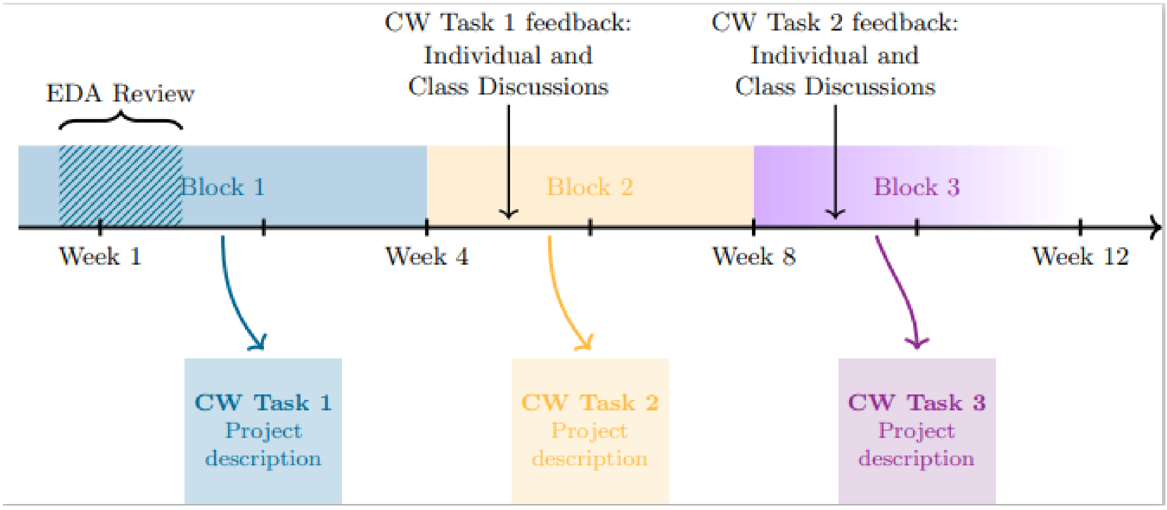

Figure 2. Assessment schedule Schematic.

Cohorts are small, with fewer than twenty students so both attention to individuals and intensity of marking is manageable. Average marks are generally higher than the average across all final year modules but typically have much greater variation and spread of grades, see table 1.

Table 1 Summary Statistics for Module Results.

Academic year	Mean	Standard deviation	LQ	Median	UQ	Max.	Min.
2021/22	71.3	11.5	62.5	68.0	75.8	94	59
2022/23	67.4	14.6	61.0	68.0	74.0	95	45

8. Reflections

It is probably unsurprising that the elements which were best executed were those we used to routinely assess. Generally analytical aspects were reasonably well understood although the computation and use of technical packages proved more problematic for some but was manageable. It should be noted that specific software was not dictated with the general advice being to use whichever was most apt for the task. This itself produced lively debate on preferences particularly with respect to ease of execution, formatting, and output quality. Students are taught RStudio, Python, and MATLAB throughout their studies and whilst some embrace all platforms others tend to adhere to a particular favourite. Conclusions in technical terms were often weak or insufficiently evidenced and model validation similarly less rigorously executed, omitted or incomplete. Typical of previous experience in similar modules.

Greater challenges were met in the more bespoke aspects of the problem-solving cycle. Crucially, the skills industry really needs when they employ ‘data scientists’, ‘data analysts’ or academic researchers.

Communicating findings in non-technical language proved to be particularly challenging. This tended to be an unsupervised section as students completed reports pre submission deadlines and notably improved across the assessments. Hence the inclusion of specific communication-based activities in further module iterations which has had some beneficial impact. Another challenging aspect is the formulation of research questions, bridging the domain-maths-domain bridge, and formulating mathematical solutions with caveats to address presented problems. Again, further exposure to example scripts has some efficacy. Approaches include extended analysis of the More or Less podcast topics, for example, identifying and accurately stipulating any statistical facts in context with reference to assumptions and limitations of the reported problem. Also, journal articles, minus abstract, are used to develop abstract writing skills, student versions are subsequently compared and further discussed alongside the author version. The latter has the added benefit of relevance to a compulsory course component, the Final Year Project dissertation.

Still, there is no doubt, mathematicians are generally very good at the mathematics of problem solving but less good at interpreting problems in mathematical terms and communicating findings, hence converting findings into recommendations for industry purposes.

In summary, I believe this to be a successful, although difficult process at undergraduate level, and am generally seeing excellent results with glowing feedback across all grade bands. Students gain demonstrable problem-solving skills, plus transferable research skills for progression to postgraduate study or the workplace.

9. References

Barrows, H.S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, 1996(68), pp.3-12. https://doi.org/10.1002/tl.37219966804

BBC (2024). *More or Less*. Available at: https://www.bbc.co.uk/programmes/b006qshd [Accessed 3 April 2024].

Bond, P. (2018). *The era of mathematics*. UK Research and Innovation.

Brewer, G. D. (1999). The challenges of interdisciplinarity. *Policy Sciences*, *32*(4), 327-337. https://doi.org/10.1023/A:1004706019826

Brewer, G. D. and Lövgren, K. (1999). The theory and practice of interdisciplinary work. *Policy Sciences*, *32*(4), 315-317. https://doi.org/10.1023/A:1004789429396

Chartier, T.P. (2015). *When life is linear: from computer graphics to bracketology*. Mathematical Association of America.

De Graaf, E. and Kolmos, A. (2003). Characteristics of problem-based learning. *International Journal of Engineering Education*, *19*(5), 657-662.

Felten, P. and Lambert, L.M. (2020). *Relationship Rich Education: How Human Connections Drive Success in College*. Johns Hopkins University Press.

Freiberger, M. and Thomas, R. (2023). *A practical guide to writing about anything for anyone!* Available at: https://plus.maths.org/content/PlusWritingGuide [Accessed 3 April 2024].

Government Open Data (n.d.). Available at: https://www.data.gov.uk/ [Accessed 3 April 2024].

Harford, T. (2020). *How to Make the World Add Up: Ten Rules for Thinking Differently About Numbers*. Hachette UK.

Krumrei-Mancuso, E.J., Haggard, M.C., LaBouff, J.P. and Rowatt, W.C. (2020). Links between intellectual humility and acquiring knowledge. *The Journal of Positive Psychology*, 15(2), pp.155-170. https://doi.org/10.1080/17439760.2019.1579359

Mathematics in Industry Reports (n.d.). Cambridge University Press. Available at: https://www.cambridge.org/engage/miir/ [Accessed 3 April 2024].

Murray, R. (2017). *How to Write a Thesis*. Maidenhead: McGraw-Hill Education.

Newman, M.J. (2005). Problem based learning: an introduction and overview of the key features of the approach. *Journal of Veterinary Medical Education*, 32(1), pp.12-20. https://doi.org/10.3138/jvme.32.1.12

Porter, T. and Schumann, K. (2018). Intellectual humility and openness to the opposing view. *Self and Identity*, 17(2), pp.139-162. https://doi.org/10.1080/15298868.2017.1361861

Shelter (2024). *Legal definition of homelessness and threatened homelessness*. Available at: https://england.shelter.org.uk/professional_resources/legal/homelessness_applications/homelessness_and_threatened_homelessness/legal_definition_of_homelessness_and_threatened_homelessness [Accessed 3 April 2024].

Spiegelhalter, D. (2019). *The Art of Statistics: Learning from Data*. Penguin UK.

V-KEMS Study Group (2022a). *V-KEMS Study Group Report: Communities of the Future*. Cambridge: Newton Gateway to Mathematics. Available at: https://gateway.newton.ac.uk/sites/default/files/asset/doc/2206/V_KEMS_Ageing_Society%20V1%20Issue%201%2031%20May%202022.pdf [Accessed 3 April 2024].

V-KEMS Study Group (2022b). *V-KEMS Study Group Report: The Public Perception of Science*. Cambridge: Newton Gateway to Mathematics. Available at: https://gateway.newton.ac.uk/sites/default/files/asset/doc/2210/The%20Public%20Perception%20of%20Science%20Virtual%20Study%20Group%20Report.pdf [Accessed 3 April 2024].

Wong, I. H. and Wong, T. T. (2021). Exploring the relationship between intellectual humility and academic performance among post-secondary students: The mediating roles of learning motivation and receptivity to feedback. *Learning and Individual Differences*, 88, article 102012. https://doi.org/10.1016/j.lindif.2021.102012

Wood, D. F. (2003). Problem based learning. *BMJ*, *326*(7384), 328-330. https://doi.org/10.1136/bmj.326.7384.328

Yew, E.H. and Goh, K. (2016). Problem-Based Learning: An Overview of its Process and Impact on Learning. *Health Professions Education*, *2*(2), pp.75-79. https://doi.org/10.1016/j.hpe.2016.01.004