Sheffield Hallam University

Embedding and assessing project based statistics

MARSHALL, Ellen

Available from Sheffield Hallam University Research Archive (SHURA) at:

http://shura.shu.ac.uk/29791/

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

MARSHALL, Ellen (2019). Embedding and assessing project based statistics. MSOR Connections, 17 (2).

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html

Sheffield Hallam University Research Archive
http://shura.shu.ac.uk

CASE STUDY

Embedding and assessing project based statistics

Ellen Marshall, Department of Engineering and Maths, Sheffield Hallam University, UK. Email: ellen.marshall@shu.ac.uk

Abstract

Traditional approaches to learning and teaching in statistics often involves the passive absorption of information through lectures, a focus on mathematical theory and assessments which test mastery of procedures. This often results in students struggling to apply their statistics knowledge in practical and authentic contexts particularly within final year projects and in the workplace. For some time, statistics educational literature has recommended shifting the focus of teaching and assessment from theory to statistical problem solving, application based statistics using real-life scenarios, and effective communication of statistics. This research has led to the production of guidelines for statistics educators from the American Statistical Association.

This paper discusses how educational literature and guidelines have been used to implement changes in the teaching of a first year probability and statistics module for mathematics undergraduates at Sheffield Hallam University. Changing to project based learning with a focus on active learning, effective decision making and communication enabled students to successfully undertake an open group project by the end of their first year. In addition, attendance, engagement and understanding were noticeably improved. The rationale, challenges and benefits to changing the focus of the course and also the teaching style are discussed.

Keywords: Statistics education, project based learning, active learning, assessment.

1. Introduction

The main goal of an introductory statistics course should be to enable students to master the basics of statistical thinking with an emphasis on application based statistics, problem based skills and effective communication of results (Cobb, 1992), which is more relevant to employment. Students are less able to reason or think statistically when the focus is on theory, methodology and mathematical calculations (Snee, 1993) and struggle to apply their statistics knowledge in a practical and authentic context. Traditional courses and exams typically test mastery of procedures and using formulae rather than the ability to apply statistical thinking and problem solving so more authentic assessment techniques need to be developed (Chance, 1997; Garfield, 1994) to test statistical thinking. Garfield, 1994, suggests that using a variety of project based assessment methods can contribute to the learning process as well as to test statistical understanding. The American statistical association guidelines for assessment and instruction in statistics education (GAISE, 2016) emphasise the importance of teaching statistics as a problem solving and decision making process rather than a collection of unrelated formulae and methods.

This paper discusses changes to the teaching and assessment within a 1st year introductory statistics course for maths undergraduates using the six recommendations of the GAISE report:

1. Teach statistical thinking;
2. Focus on conceptual understanding;
3. Integrate real data with a context and a purpose;
4. Foster active learning;
5. Use technology to explore concepts and analyse data;
6. Use assessments to improve and evaluate student learning.

2. Methods

2.1. Teaching statistical thinking

The process of the statistical problem solving approach (Marriott et al., 2009) shown in Figure 1 was embedded from the start of the module. The process starts with a motivating research question, considers suitable data collection or manipulation and the appropriate techniques to address the research question. Students on this course learnt to use Excel or SAS to carry out suitable analysis in addition to the mathematical calculations and how to report the results effectively.

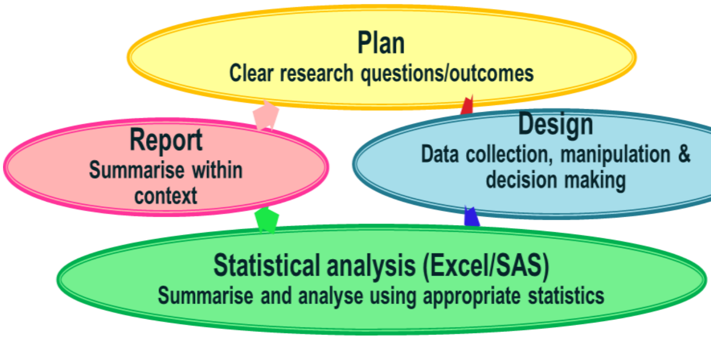


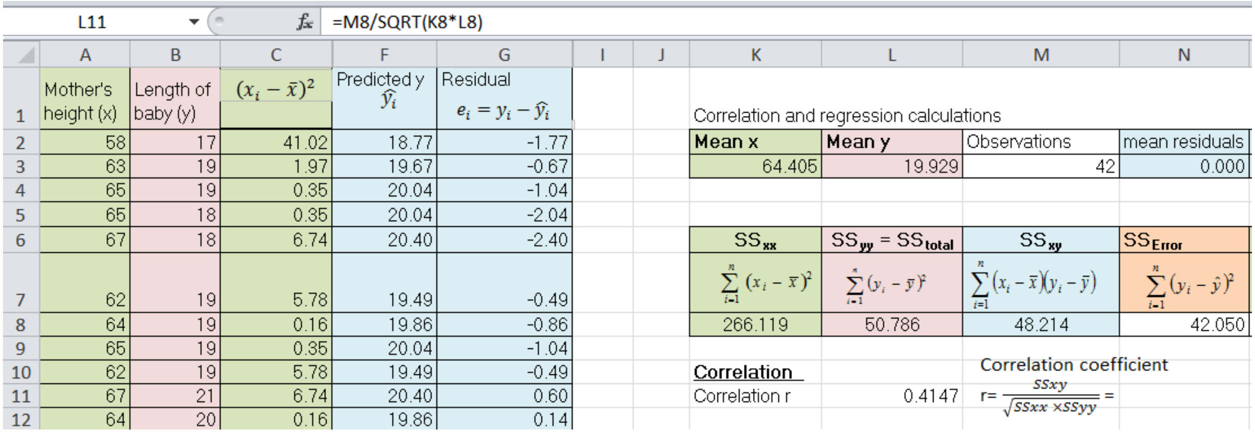
Figure 1: Statistical process cycle

Only simple summary statistics and charts were used in initial teaching and assessment which enabled students to master the terminology and the process for statistical problem solving before moving on to hypothesis testing. The assessments included individual reports and a group presentation which tested all aspects of the statistical process cycle including the selection of appropriate techniques.

2.2. Conceptual understanding and the use of technology

In the past, the content of statistics modules was limited to what could be computed mathematically and very little has changed in many institutions despite the advances in computing power and how statistics is used in the workplace. The general focus of statistics educational research centres on shifting the focus to conceptual understanding over mathematical procedures, (Garfield, 1995; Cobb, 1992, Snee, 1993). In addition, students generally find the core concepts of statistics difficult and need more time to develop conceptual understanding (Tishkovskaya, 2012; Garfield, 1995). To build in more time for conceptual understanding, the guidelines suggest reducing the time spent on probability theory, summary statistics, using the critical value method of hypothesis testing and large numbers of procedural calculations which do not enhance the students' conceptual understanding. The use of technology enables students to speed up calculations, carry out more complex tests and visualise concepts such as variability between sample means.

Students are taught to use SAS to perform statistical analysis, as many employers of our graduates use this package, and the maths taught alongside enables students to better understand the output. Excel is also used as a tool for speeding up mathematical procedures, demonstrating the impact of changes to data and to allow students to convert their mathematical understanding in a more practical way. For example, Excel is used to demonstrate the impact of outliers on means and standard deviations by changing one value to an outlier and to produce templates for statistical techniques. To improve their understanding of regression, students were asked to produce a spreadsheet for calculating regression equations and checking assumptions (see Figure 2 below). This demonstrated their understanding of the mathematical calculations without large amounts of time being wasted on calculator calculations in class or in an exam. Students reported that this was a useful exercise.



L11 | =M8/SQRT(K8*L8)

	A	B	C	F	G
1	Mother's height (x)	Length of baby (y)	$(x_i - \bar{x})^2$	Predicted y $\hat{y}_i$	Residual $e_i = y_i - \hat{y}_i$
2	58	17	41.02	18.77	-1.77
3	63	19	1.97	19.67	-0.67
4	65	19	0.35	20.04	-1.04
5	65	18	0.35	20.04	-2.04
6	67	18	6.74	20.40	-2.40
7	62	19	5.78	19.49	-0.49
8	64	19	0.16	19.86	-0.86
9	65	19	0.35	20.04	-1.04
10	62	19	5.78	19.49	-0.49
11	67	21	6.74	20.40	0.60
12	64	20	0.16	19.86	0.14

Correlation and regression calculations

K	L	M	N
Mean x	Mean y	Observations	mean residuals
64.405	19.929	42	0.000
SS_{xx}	$SS_{yy} = SS_{total}$	SS_{xy}	SS_{Error}
$\sum_{i=1}^{n}(x_i - \bar{x})^2$	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$	$\sum_{i=1}^{n}(y_i - \hat{y})^2$
266.119	50.786	48.214	42.050
Correlation		Correlation coefficient	
Correlation r	0.4147	$r = \frac{SSxy}{\sqrt{SSxx \times SSyy}} =$	

Figure 2: Using Excel to understand regression

Research suggests that the use of simulations in statistics education can be effective in improving student's statistical reasoning (Zieffler et al, 2008) and understanding of the core concepts of hypothesis testing. Students were given different samples of data to aide in their understanding but next year, Excel templates will be used to generate random samples and demonstrate distributions e.g. changing the parameters of the normal to observe the change in shape.

Although the focus of the module was on project based statistics, the number of formulae covered still resulted in the students feeling overloaded. In the next academic year, some of Excel and data summary aspects of the statistics course will be covered in a 'Maths Technology' module and less time will be spent on Probability so that the core concepts can be covered more slowly, in line with the suggested guidelines for reducing content.

2.3. Integrating real data with a context and purpose

Most HE students encounter statistics as part of their course or research but negative attitudes and anxieties can contribute to difficulties grasping the basic concepts of probability and statistics (Gal and Ginsburg, 1994). Using interesting research, media articles and data sets can make the learning process more positive (Zieffler et al, 2008; Chance, 1997), and demonstrate the relevance of course material in everyday life (Neumann, Hood and Neumann, 2013). Most statistics courses use different data for almost every tutorial question and often variables are referred to as x and y which doesn't help a student see the whole statistical process or how examples relate to real life. Where possible, this module used a few large key datasets to link the different parts of the course so that students could observe the whole process and start to see statistics as a set of related rather than unrelated techniques. .

One data set used is the Titanic data set which can be linked to several newspaper articles including the one in Figure 3 and allows a range of research questions centering on which variables impacted on survival. The introduction of this data set prompted more noticeable engagement within the lecture and allowed students to investigate claims made by papers with real data including whether or not we can ascertain that more British people died because they queued.

Which variables could you use to investigate this statement made in the newspaper article below and which summary statistics could be used?


INDEPENDENT News InFact Politics Voices Indy/Life Business

News › World › Australasia

More Britons than Americans died on Titanic 'because they queued'

Kathy Marks | @kathymarksoz | Wednesday 21 January 2009 00:00 GMT | 0 comments

Like Click to follow The Independent Online

British passengers on the Titanic died in disproportionate numbers because they queued politely for lifeboats while Americans elbowed their way on, an Australian researcher believes.

Figure 3: Example of newspaper discussion topic

The Titanic data set which is primarily used for categorical analysis and the Birthweight data set used for correlation and regression are both available via the Sheffield Hallam maths support webpages along with examples of using these data sets to demonstrate statistical techniques. (See: https://maths.shu.ac.uk/mathshelp/Resources/index.html).

At the beginning of the course students had the opportunity to undertake a voluntary ‘Data Challenge’ which involved searching for data and articles which were more interesting than the lecturers’ examples. Surprisingly, despite not being assessed or widely advertised, about a 3rd of students participated with the winners being decided by votes and the views of the statistics team. A summary of the articles and data, which could be used by lecturers within introductory courses, can be obtained from the author.

2.4. Fostering active learning

Many studies have shown that problem based learning using real life data and scenarios improves statistical thinking and engagement particularly when engaging with real data and scenarios (Marriot et al., 2009; Rossman et al, 2006; Garfield and Ben-Zvi, 2007; Garfield, 1995) and this motivation leads to better learning outcomes. Statistics educational literature also suggests that using active learning within the class is more effective than the passive absorption of traditional lecturing (Cobb, 1992; Freeman et al, 2014). Active learning is an aspect of constructivist teaching where students construct their own meaning by building on from existing knowledge and through the use of authentic tasks. Methods for active learning include problem based learning through case studies, simulation, the use of technology generally, cooperative or collaborative learning. Embedding brief activities to encourage students to think about statistical interpretation and understanding before being told the answer or method (Tishkovskaya, 2012; GAISE) is a simple way of introducing active learning within the classroom. Many useful teaching resources and explanations of different types of educational methods can be found on the CAUSEWEB (https://www.causeweb.org/cause) and STATS 101 toolkit (https://community.amstat.org/stats101/home) sites.

Prior to the author teaching on the module, attendance and engagement with statistics were very low. Students were given a full set of notes and many did not see the point in attending resulting in poor understanding of the topics. The author used simple applications of active learning such as gapped notes, interactive lectures and encouragement of collaborative learning to successfully improve engagement, attendance and understanding. The gapped notes included questions to be answered from the lecture material presented, rather than gaps to copy exactly what was on the board, to ensure students understood the lecture material. The interactive lecture activities were designed to either get students to think statistically before explaining a concept, interpret results or to apply what they have learnt. Students were able to construct their own meaning by building up from knowledge of more basic techniques which also allowed them to connect the different aspects of statistics and deepen their understanding. In addition to the benefits of breaking up lectures for students, this also enabled the lecturer to assess understanding by the ease at which students undertook the activities and questions being asked. Students said that they preferred the gapped notes to a more traditional lecture as it kept them engaged and enabled them to reflect on their understanding.

Students had access to written booklets and lecture slides but the average student attended 78% of statistics lectures which was a considerable improvement on the module attendance prior to the changes. It is well known that the relationship between attendance and performance is usually weak – moderate but Figure 4 shows a strong correlation (r = 0.7) between lecture attendance and overall coursework mark for this module which suggests that attending the lectures was beneficial to learning. It should be noted that attendance relates purely to the statistics lecture as attendance at tutorials and the probability lectures was not consistently recorded. The regression model suggests an improvement of approximately 6% on coursework for each 10% increase in attendance (approximately two lectures) and that 50% of the variation in performance can be explained by attendance. Students were generally positive about the style of the teaching and when asked to give advice to new students on the standard University module feedback form, the most common suggestions were "attend all the classes" and "fill in all the booklets which are really useful".

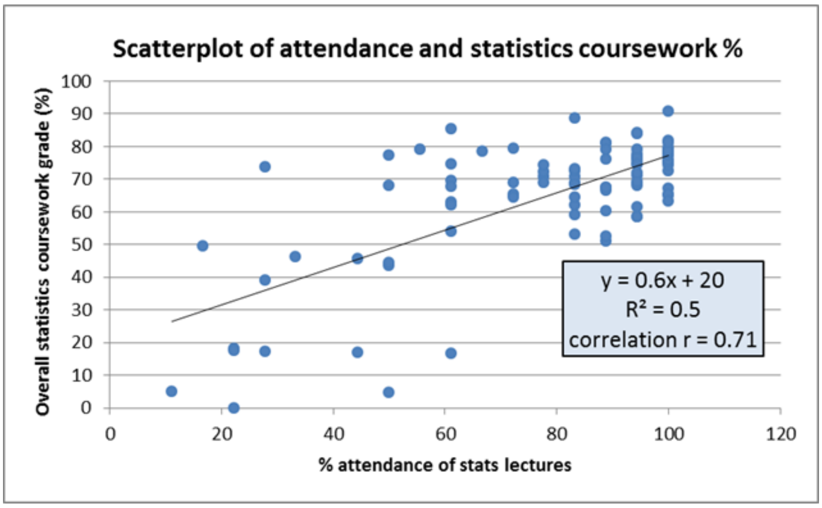


Figure 4: Scatterplot showing the relationship between attendance and performance

2.5. Using assessment to improve and evaluate learning

Statistics when taught to maths undergraduates has traditionally focused on the mathematical procedures of statistics and students learn to recognise when to use given formulae to compute final answers. These processes are usually tested through exams or closed questions in coursework with the emphasis on the correct answer rather than the interpretation. This reinforces the view that statistics is a set of unrelated mathematical topics and does not develop statistical thinking. Garfield (1994) sees assessment of problem based skills as an integral part of teaching and learning of statistics and adds that as students tend to value what is assessed, assessments testing statistical thinking and communication should be introduced. These aspects were tested through two written reports and a group presentation which this year all used the same data set. Some students were initially sceptical about the 'non-maths' aspects of the course but the use of the diagram in Figure 5 to explain the different skills they would need in the workplace helped them understand and appreciate the rationale behind the teaching and assessment.

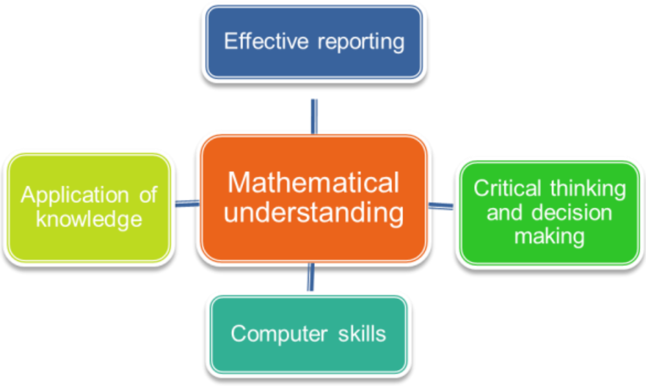


Figure 5: Skills needed by mathematician/statistician

They also struggled with the first project based assignment as it was different to the traditional maths assessment but by the last assignment they had a greater appreciation of what statistics really is and even found the collaborative scenario based presentation enjoyable.

In the first assessment, students were given a number of research questions and a dataset from a medical trial. Students were expected to select appropriate summary statistics and graphs for each question and communicate these results efficiently and effectively in the form a report. The understanding of data types and selection of appropriate statistics to address specific hypotheses was tested trough an online test and the visual representation, calculations and interpretation were assessed through a formal written report. Students generally performed well on the online test but did not necessarily connect the questions from the test to what was expected to be included within the more open report. The negatives were that the marking of the reports was very time consuming and that the students didn't feel that they had been adequately prepared for the writing of the report. In the next academic year, students will receive guidance on report writing; the assessment criteria will be more clearly explained and they will be given the opportunity to get feedback on formative sample questions within tutorials. To ease marking, only one individual written report will be assessed and two parts will be submitted; a formal report and an Appendix containing calculations and SAS output.

The final assessment, a cooperative group presentation involving analysis of data from a medical trial, enabled students to fully express statistical reasoning, thinking and communication in a more open project. Students seemed genuinely enthusiastic about working on a real life scenario and commented on how their understanding had improved through the cooperative learning. The team of staff assessing the twenty six presentations were impressed by the students' ability to apply statistical thinking and a wide range of techniques appropriately. Next year the students will be undertaking work related collaborative assessment for external companies who will provide data and ideas for statistical analysis and it is hoped that this will engage students further with statistics.

2.6. Reflections and feedback

The author's primary role is in statistics support and has subsequently observed the gaps in student knowledge when undertaking their own research. One of the main aims of the author was to integrate statistical skills particularly choosing suitable techniques and effective reporting into the first year module to adequately prepare the students with roles within industry alongside the traditional teaching. Whilst the author had some material, data and examples from other courses and notes from her predecessor, making fundamental changes to the structure, aims and content of any module is time consuming and ongoing particularly with a module that students traditionally dislike. Staff who teach on further stats modules were impressed with the understanding and statistical skills demonstrated by this cohort of students when watching the group presentations. They also observed a noticeable improvement in student understanding when the students started their 2nd year stats module.

Students currently in their 2nd year were asked to comment on the course and were very positive about the structure, content and style of teaching in general and still use the course booklets to help with their understanding of further statistics modules. They particularly liked the effective use of gapped notes and exercises within the lectures as a means of staying engaged and strengthening understanding. Many students started the course with negative views about statistics based on their A level experiences but a number of these students have now expressed an interest in careers in statistics and have applied for statistics related placements.

3. Conclusion

For many years, statistics educational literature and guidelines have suggested that a shift in focus from mathematical calculations to statistical problem solving and communication is needed to meet the research and employment needs of students of today. This paper discusses changes made to an introductory probability and statistics course to ensure students are able to understand and apply statistics in a real world setting and how the current guidelines and suggestions for statistics education (GAISE) can be implemented. Introducing and reinforcing the problem solving approach throughout a course, using large data sets and real life scenarios, using active learning within lectures, using technology to enhance learning and changing assessments to test decision making and communication were all successfully implemented within a first year statistics module which led to improved understanding and engagement. Making substantial changes to a module is an iterative process however and further adaptions to the teaching will be made in the next academic year including more simulation and constructivist teaching allowing students to create rather than memorise formulae.

4. References

American Statistical Association, 2016. Guidelines for the assessment and instruction of statistics education. Available at: http://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx [Accessed 11 March 2019].

Chance, B., 1997. Experiences with Authentic Assessment Techniques in an Introductory Statistics Course, *Journal of Statistics Education, 5(3).* Available at: http://www.amstat.org/publications/jse/v5n3/chance.html [Accessed 11 March 2019].

Cobb, G., 1992. Teaching Statistics. *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. Steen. MAA Notes, No. 22, Washington: Mathematical Association of America, pp.3-34.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., and Wenderoth, M.P., 2014. Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences of the United States of America,* 111, pp.8410–8415.

Gal, I., and Ginsburg, L., 1994. The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework. *Journal of Statistics Education, 2(2)*. Available at: http://www.amstat.org/publications/jse/v2n2/gal.html [Accessed 11 March 2019].

Garfield, J., 1994. Beyond Testing and Grading: Using Assessment to Improve Student Learning, *Journal of Statistics Education, 2(1)*. Available at: http://www.amstat.org/publications/jse/v2n1/garfield.html [Accessed 11 March 2019].

Garfield, J., 1995. How Students Learn Statistics, *International Statistical Review, 63(1),* pp.25-34.

Garfield, J., and Ben-Zvi, B., 2007. How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75(3),* pp.372-396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Marriott, J., Davies, N., and Gibson, L., 2009. Teaching, Learning and Assessing Statistical Problem Solving, *Journal of Statistics Education, 17(1)*. Available at: http://www.amstat.org/publications/jse/v17n1/marriott.html [Accessed 11 March 2019].

Neumann, D., Hood, M., and Neumann, M., 2013. Using Real-Life Data when Teaching Statistics: Student Perceptions of this Strategy in an Introductory Statistics Course. *Statistics Education Research Journal,* 12, 59-70. Available at: http://iaseweb. org/documents/SERJ/SERJ12%282%29_Neumann.pdf [Accessed 11 March 2019].

Snee, R., 1993. What's missing in statistical education? The American Statistician, 47(2), pp.149-154.

Tishkovskaya, S. and Lancaster, G.A., 2012. Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education,* 20(2).

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K. and Chang, B., 2008. What does research suggest about the teaching and learning of introductory statistics at college level? A review of the literature. *Journal of Statistics Education.* 16(2). https://doi.org/10.1080/10691898.2008.11889566