Bring Your Own Dataset: A Student-centred Approach to Teaching Biostatistics

Dr Agus Salim & Assoc Prof Gerald Koh Choon Huat
Saw Swee Hock School of Public Health

Background & Introduction

Biostatistics is broadly defined as the application of statistical techniques to scientific research in health-related fields, including medicine, biology, and public health, and the development of new tools to study these areas. Biostatistics plays an integral part in public health research and is used as one of the main analytical tools in areas such as identifying disease risk factors, determining the outcome of public health interventions, and forecasting trends of disease burden. In recognition of this, most graduate programmes in public health require students to take one or more Biostatistics modules that would equip them in their research and professional career. In our Master of Public Health (MPH) programme, we offer two Biostatistics modules and the first author has been teaching one of them, COS5218 "Advanced Quantitative Epidemiology" (AQEM), which is an elective module taken by students with clinical epidemiology specialisation. The module has always been taught as a 1-week teaching block where each day is dedicated to the teaching of one of the following topics:

- multiple linear regression,
- logistic regression,
- survival analysis, and
- analysis of repeated measurements

The four topics cover different major statistical methods that the students will most likely need to use in their professional career. The first author co-teaches this module with another colleague and each teach two topics. In his case, he teaches logistic regression and the analysis of repeated measurements.

In the AQEM module, students are exposed to practical sessions using STATA software (*www.stata.com*). The practical session is meant to equip students with experience in analysing real datasets in order to tackle real-life research questions.

In a typical practical session, students use a real dataset that we have prepared. In the first part of the session, the lecturers guide the students to apply the statistical methods they learnt during the lecture. In the second part, the students have to complete an exercise in which they need to use appropriate statistical methods and certain variables in the dataset to answer a series of scientific questions that we have prepared. The students' answers to this exercise form the basis of their individual reports and group presentations which together constitute 50% of the total assessment for this module. While students are required to attend practical sessions for all four topics, for the purpose of the group presentations and individual reports, students are randomly allocated to one of the four groups, with each group being responsible for one topic.

Recommended Citation

Salim, A. & Koh, G.C.H. (2013). Bring your own dataset: A student-centred approach to teaching biostatistics. *CDTL Brief, 16*(1), pp. 17-21.

Conducting the Proposed Scheme

Over the years, through personal observations and informal chats with students, it was noticed that their interest in the practical session varied greatly from one student to another. Based on the information gathered from such chats, the main reason why some students were not interested in the exercise was because the dataset (prepared by the lecturers) often comes from a branch of medicine/health science which is different from the student's own professional background. This lack of interest is of particular concern as students enrolled in the MPH come from various educational and work backgrounds. A student from a gastroenterology background, for example, may not be very motivated to answer exercise questions related to a dataset that comes from a cardiovascular research setting. This is because the student does not see the relevance of such questions to his/her current and future work. This perception could not only affect the student' interest and motivation for the entire course, but it could also affect his/her view on the usefulness of biostatistics for his/her research and professional career. Some articles have argued that a student-centred approach to learning, where students set their own goals and resources to achieve those goals (Jonassen, 2000), was more meaningful to them (Pedersen & Liu, 2003).

To provide MPH students with a better learning experience, the first author proposed an alternative approach where he got them to do *authentic activities* (Herrington et al., 2003). Authentic activities refer to learning activities where students face problems that closely resemble those that real-life professionals have to deal with. Such activities have been used successfully across many disciplines. Examples of authentic activities include the design and construction of a Formula SAE[1] race car as part of a mechanical and mechatronics engineering curriculum (Bullen & Karri, 2002). Hunt et al (2002) also described a project which was a good example of an authentic activity, where instead of the more conventional orientation activities, students addressed issues related to the transition from high school to university by shooting videos of their peers exploring the university campus and capturing on film the nature of university life. Similarly, Marshall et al (2001) used authentic activities to design a course teaching mathematics to indigenous adults.

In our proposed authentic activity, students are allowed to bring their own dataset to the practical session and with the first author's input, form exercise questions out of their own dataset (see details below). This activity was designed to mimic the situation students would face when they become public health researchers where they would have to perform data analysis and interpret the results.

Methodology

To give students enough time to decide, an email was sent five weeks before the first lesson to all 12 students that have registered for the AQEM module in Academic Year (AY) 2010/2011. The email invited them to participate in this alternative teaching scheme where they can bring their own dataset. The potential benefits of participating in the scheme were outlined which includes working on exercise problems that are more relevant to their research interests, with the possibility of developing the individual report further into a research paper if the student can come up with sufficiently novel and relevant scientific questions. Out of 12 students, four replied to confirm their interest. These students were invited to bring their dataset in for a preliminary one-on-one meeting with the first author.

At the preliminary meeting, the suitability of the dataset was assessed with input from the student in terms of:

1. the complexity of the variables in the dataset,
2. the lack of potential issues that may distract students (such as a large amount of missing data), and
3. whether the two statistical methods the first author was teaching (logistic regression and analysis of repeated measurements) were appropriate to answer the research questions from the dataset.

The first two aspects of the assessment are needed to ensure that the student's dataset has enough complexity so that the exercise questions set were of comparable levels of difficulty to the exercise questions from the original dataset prepared by the lecturers. The third aspect of assessment is needed because the first author could only assess presentations and reports of those students that used statistical methods under the two topics that he was teaching.

The first author spent about an hour with each student, discussing and assessing the appropriateness of their dataset. All four students expressed great interest in the scheme but after going through the assessment with each of them, it was decided that one of the student's dataset was not appropriate because the research questions that were put forward could be more appropriately answered using statistical methods that fall under the topic of linear regression (which the other co-lecturer taught). The remaining three students were invited for a second one-on-one meeting in which the first author and the student worked together to set the exercise questions. The students were allowed to come up with their own research questions while the first author facilitated the process by giving input to ensure that certain standards are maintained, specifically that the questions they set were of a comparable level of difficulty to the questions prepared by the lecturers. To maintain objectivity while grading, the first author tried to be on the 'sidelines' as much as possible by letting the students came up with their own questions and by purposely not examining the dataset beyond checking its basic characteristics.

Feedback

In the post-module feedback, all three students were very positive about the alternative scheme and all of them would recommend that other students participate in it. One student noted that being able to use their own dataset means that the practical session has now "…become particularly relevant and applicable as the student took [a] personal interest in it".

Another student commented on how the scheme was helpful in building his confidence in performing data analysis ("…it allows me to analyse data with more confidence") and how the preliminary discussion meetings with the lecturer can potentially open avenues for research collaboration between the lecturer and the student ("…possible for potential collaboration in research").

Conclusion

Although conducting this scheme was time-consuming, we feel that it is worth the investment of time, as it certainly improves students' confidence and learning experience, as evidenced from their feedback. One of the students continued to collaborate with the first author even after completing the course. Since then, they have submitted a paper to a peer-reviewed journal. The content of that paper was partially based on the research questions they set for the exercise.

However, because of the sheer amount of preparation time the lecturer needs to invest in, which includes assessing the dataset and facilitating the process of setting separate exercise questions for each student, this kind of scheme cannot be offered to a large class. However, it can still be implemented in a medium-sized class (< 50 students) if some students are willing to share their datasets with others.

There is also the issue of comparability of questions for grading purposes. Throughout the scheme, great effort was expended to ensure that the questions from the different datasets were of comparable levels of difficulty. To do this, the first author sometimes had to convince a student to tackle a question that he/she perceived as being more difficult. In such cases, he usually tried to demonstrate to the student the potential benefits of attempting the more 'difficult' analysis and how it can potentially provide better insights, instead of taking the 'easier' option. For the AQEM class, the marks the three students received do not indicate

continued on the next page...

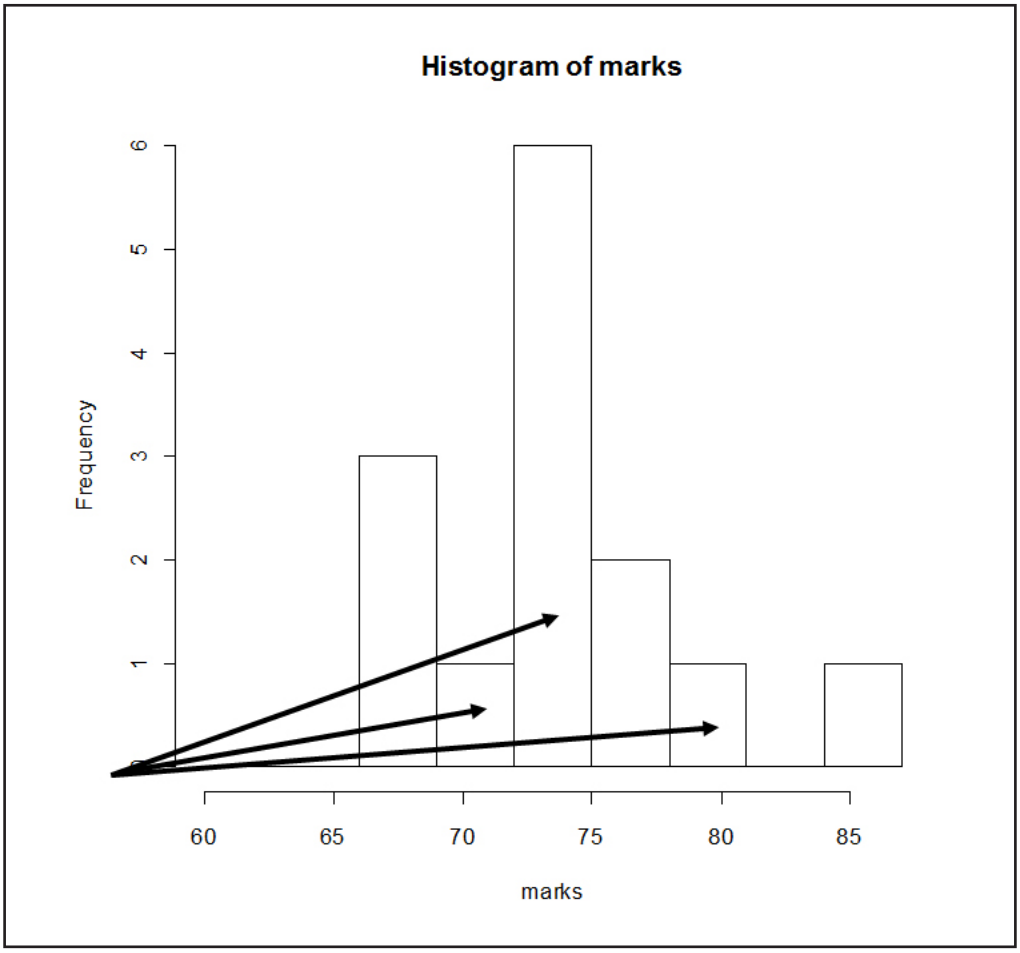

Figure 1. Distribution of the final grades for all students. The arrows indicate the intervals which contain grades for the three students participating in the scheme.

that they were at a disadvantage for using their own datasets, compared to their classmates who used datasets provided by the lecturers (see Figure 1). Nevertheless, we feel that these challenges should not prevent the running of this teaching strategy again for subsequent cohorts.

About the Authors

Dr. Agus Salim taught several graduate Biostatistics modules at the Saw Swee Hock School of Public Health (SSHSPH). His lifelong teaching goal is to make the learning of statistics accessible and fun for students without prior statistics or mathematics background.

Assoc Prof Gerald Koh currently teaches family medicine, geriatric medicine and public health to graduate students at SSHSPH. He received both the Faculty Teaching Excellence Award and the University Teaching Excellence Award in 2009.

Endnote

1. "SAE" stands for Society of Automotive Engineers.

References

Bullen, F. & Karri, V. (2002). Design and construction of a Formula SAE racecar in a teaching and research framework. In A. Goody, J. Herrington & M. Northcote (Eds.), *Quality Conversations: Research and Development in Higher Education,* Volume 25 (pp. 74-82). Jamison, ACT: HERDSA. [verified 6 Feb 2003] http://www.ecu.edu.au/conferences/herdsa/papers/ref/pdf/Bullen.pdf.

Herrington, J., Oliver, R. & Reeves T.C. (2003). Patterns of engagement in authentic online learning environments. *Australian Journal of Educational Technology, 19*(1), pp. 59-71.

Hunt, L., Kershaw, L. & Seddon, J. (2002). Authentic transitions: The click around ECU online transition to university program. In A. Goody, J. Herrington & M. Northcote (Eds), *Quality Conversations: Research and Development in Higher Education,* Volume 25 (pp. 338-353). Jamison, ACT: HERDSA. [verified 6 Feb 2003] http://www.ecu.edu.au/conferences/herdsa/papers/ref/pdf/Hunt.pdf.

Jonassen, D.H. (2000). Revisiting activity theory as a framework for designing student-centered learning environments. In D.H. Jonassen & S.M. Land (Eds.),*Theoretical Foundations of Learning Environments* (pp.89–121). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Marshall, L., Northcote, M. & Lenoy, M. (2001). Design influences in the creation of an online mathematics unit for indigenous adults. In Kennedy, G., Keppell, M., McNaught, C. & Petrovic, T. (Eds), *Meeting at the Crossroads. Short Paper Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education* (pp. 113-116). Melbourne: Biomedical Multimedia Unit, University of Melbourne. http://www.ascilite.org.au/conferences/melbourne01/pdf/papers/marshalll.pdf.

Pedersen, S. & Liu M. (2003). Teachers' beliefs about issues in the implementation of a student-centred learning environment. *Educational Technology Research and Development, 51*(2), pp. 57-76. ■