PROBLEM BASED LEARNING IN A BIOSTATISTICS COURSE

J.P. Cruz

University of Aveiro (PORTUGAL)

Abstract

We have introduced statistical problems, to be solved using the R software, into a Biostatistics course, in order to increase motivation for the field that requires a certain level of mathematical knowledge when most students are not always inspired for it. Our traditional class style used to be based only on slide presentations followed by pen and paper exercises with a calculator. Our aim was to complement this method with the use of software as a professional tool creating a active learning environment. Students came from Biology degree, Teaching of Geology and Biology degree and Marine Sciences degree.

Each of the 200 students were presented with a total of four problems, during the semester, in the topics of Descriptive Statistics, Inference in One Variable, ANOVA and Simple Linear Regression. Students were requested to solve them at home and answer them in a form available in the "Moodle Inquiry" tool. Each student has his own different sample and also, questions were parameterized. For example, questions about Confidence Intervals were posed with different confidence levels (90%, 95% or 99%). Each students sees a different problem. Each of these has more than ten parameterized questions related to the same dataset exposed in the beginning of the text. Moodle doesn't do this type of deliver different composed problems to each student so a small Python library was used to generate different problems and evaluate each individual student answer (numerical, textual or multiple choice types).

To evaluate our methodology, we request students to "Share ideas, thoughts and constructive judgments about the Problems and also about the course" while students were working in the third Problem and also after the First Written Evaluation. The last and fourth Problem has been answered in class and students were requested to grade sentences in a five item Likert scale. Questions were about effort, time, help from other students and help from teacher.

The analysis of answers suggest that the methodology of Problem Solving should be used again, with improvements, given the motivation and enthusiasm it promotes.

Keywords: problem based learning, statistics

1 INTRODUCTION

Problem based learning [1] is an active learning technique that has been introduced into a Biostatistics course at University of Aveiro since the start of the subject a decade ago. By that time, fewer students were enrolled and the PBL approach was as follows: students have to find their own research problems meeting professional researchers at their departments. This methodology use to work very well and motivate students. However, the increased number of students and almost the same number of researchers and contexts had run out new problems. The problem arose when many students started to modify previous works and the motivatoin has declined. They start to complain of the scarcity of resources to produce new studies next to researchers in their fields. Another issue relates with the statistical topics that use to be only one of two: Linear Regression or ANOVA subjects, leaving other matters of the curriculum out of scope of the PBL.

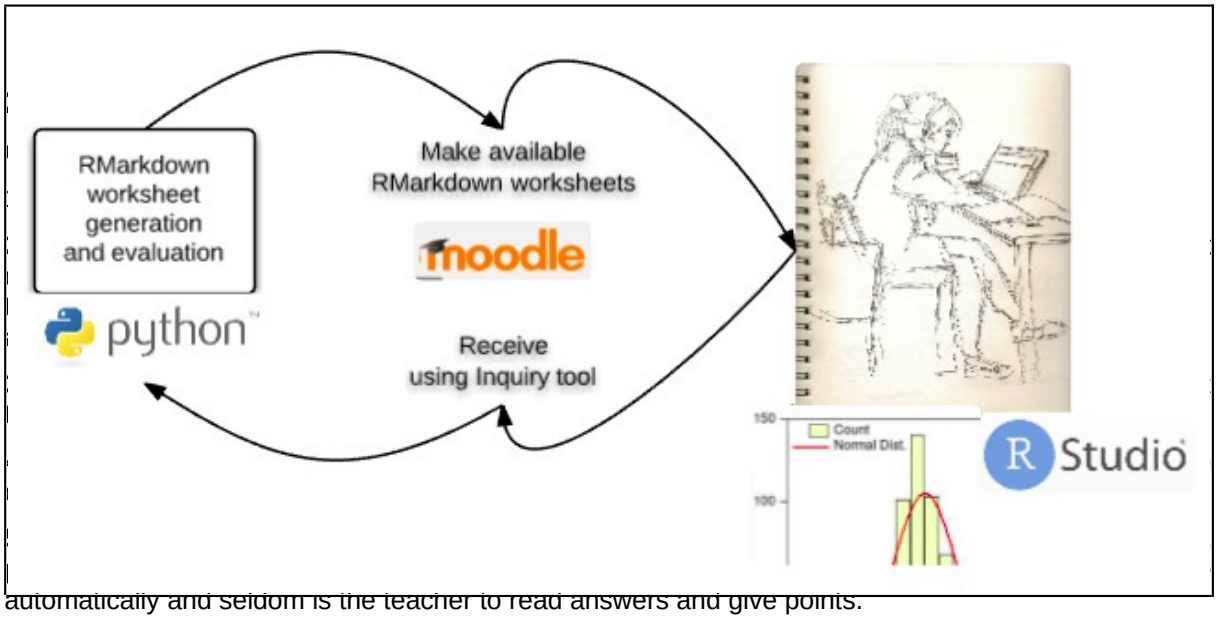
To improve the above aspects we have proposed four problems in four moments of the semester, covering Descriptive Statistics, Statistical Inference (confidence intervals and statistical hypotheses), ANOVA and Linear Regression.

The software chosen to reshape the PBL approach was using the R language and system [2] and RStudio [3] as a powerful environment: while, as said, "R is a language and environment for statistical computing and graphics", RStudio is an open-source integrated development environment (IDE) for R with reporting capabilities, that were used in by students, and many other tools for professional statisticians and data analysts. RMarkdown [4] is an utility of RStudio to produce documents that are fully reproducible and is a form of 'literate programming' because it is easier for

humans to read and understand (Knuth, 1984). It allows putting together text and code to produce reports without leaving the tool (avoiding exports and copy/paste to other tools).

In our usage of PBL, each RMarkdown worksheet contains a different statistical problem with different datasets for each student, as well as different parameterized questions. One worksheet for each of the four mentioned topics are given to the student to be solved at home.

Communication is done using the local campus Moodle. However, up to our knowledge, Moodle does not have a parameterized generator for 'projects', i.e., a long set of parameterized questions related to the same context at the start of a problem. Therefore, the generation have been done outside Moodle, using a small Python [5] script producing a RMarkdown worksheet for each student. On their side, students use Moodle 'Inquiry tool' to answer all the questions in their problems. The teacher collects answers and the Python script generates reports with grades and solutions to each student. Fig. 1 depicts the process.



automatically and seldom is the teacher to read answers and give points.

To evaluate the methodology an questionnaire was done using Moodle. Results will be summarized in the Section 3.

2.1 Some details of the structure

Each worksheet contains: (a) the name of the student; (b) a textual description of a scenario within a certain statistical context; (c) a different data set for each student; (d) R commands that produce graphical plots, tables or values; (e) parameterized questions about those R results. Only seldom, the student is requested to introduce

Each question in the worksheet has an equivalent in the Moodle Inquiry Tool. Student has to copy/paste his answer to Moodle for each answer. Sometimes, the Inquiry Tool has more questions than the original worksheet.

After the deadline, all results are collected in a spreadsheet and a Python script compares all results with the appropriate true answers to calculate the grade. In this process, prepared comments are introduced back to the worksheet as if a teacher was marking. Fig. 2 shows a sample.

Question: Do you observe a linear relation of the advance of the costal line into the sea?
Your answer: "yes, there is a linear relation"
Correct answer: "there is a linear relation"
Your answer is: correct
Suggestion: The variability on the position of the costal line is due to the some important cause and only a small part are due to random effects.

Figure 2 - Sample of the report the student sees after correction

2.2 What are the students saying

In the middle of the semester, an open answer questionnaire was given to students with the following text (the original was in portuguese):

"Share ideas, thoughts, positive or negative judgments about the techique of Problems and also about the UC. Present also self-evaluation while a Biostatistics student. Thank you."

The analysis of answers concerning the PBL follows.

3 RESULTS

Our Bioestatics course have, approximatly 200 students enrolled, 140 choose to work on given problems of our PBL approach. From these, 90 students decided to answer the question posed on section 2.2. We have extracted cases relevant to the PBL procedure and Table 1 summarizes the results.

Table 1. Empirical analysis of open answers

Summary of main opinions about PBL approach	Freq.	Perc.
Helps schedule study and autonomous study	24	53%
A dedicated class must be given to explain software	6	13%
Makes study an amusement and dynamic	4	9%
Works must be done in Excel and not in RStudio	4	9%
Technical difficulties with R are hard	3	7%
Current (PBL) method is better that previous method	2	4.5%
Some question Problems are too difficult	2	4.5%
Total of opinions with respect to PBL usage:	45	100%

From the summary, and from level of preferences, we see that the PBL approach should continue with changes. It is assured that it helps study times, and possible in an strongly motivated way, by looking at "Helps schedule study and autonomous study" (54%), "Makes study an amusement and dynamic" (9%). Of course this values must increase with a better approach but consideration must be taken that the question was open (not all students remember to enphasize these aspects).

Most of the students that did not pass the course in previou year decided not to enroll in PBL because they already have grades from previous works. The vast majority of students did not knew the previous procedure and that's why only 4.5% say it better this new approach.

The other opinions like "A dedicated class must be given to explain software", "Works must be done in Excel and not in Rstudio", "Technical difficulties with R are hard" and "Some question Problems are too difficult" are calling for improving the procedure.

4 CONCLUSIONS

We have presented an approach to PBL giving RStudio worksheets to students, each student with a different worksheet to be worked at home. The worksheet contains an extensive set of questions related to the same dataset simulating a real world example. This approach have been appreciated bu students has it make them study more regularly and more motivated. However, improvments are to be planned in order to increase the level of appreciation of the method.

A small script of Python was made to deliver different in data and parameterized questions inside Rmarkdown worksheets. This script provides students with a report of their sucesses and failures in each question and should be operate more easily or integrated elsewhere.

ACKNOWLEDGEMENTS

This seminar was supported in part by the Portuguese Foundation for Science and Technology (FCT – Fundação para a Ciência e a Tecnologia), through CIDMA - Center for Research and Development in Mathematics and Applications, within project UID/MAT/04106/2019.

REFERENCES

[1] M. Savin-Baden, Problem-Based Learning in Higher Education: Untold Stories. Biochem. Mol. Biol. Educ. Biochemistry and Molecular Biology Education. 31, 454- (2000).

[2] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[3] RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

[4] Dvorak, T., Halliday, S. D., O'Hara, M., & Swoboda, A. (2019). Efficient empiricism: Streamlining teaching, research, and learning in empirical courses. Journal of Economic Education, 50(3), 242-257.

[5] Python Software Foundation. Python Language Reference, version 3.6. Available at http://www.python.org.