

ONLINE DOCUMENTATION FOR A BIostatISTICS COURSE USING MODERN TECHNOLOGY

J.P. Cruz, A. Freitas

University of Aveiro (PORTUGAL)

Abstract

In the last three years, the first author has been producing an online documentation to be applied on a Biostatistics course in a public university, to make the study of this subject more dynamic. The target students come from Biology, Geology and Biology, and Marine Sciences degrees. The main goal is to increase the study motivation outside classes offering better study support and providing a means to share new explanations online easily and quickly about subjects when certain frequently asked questions arise. This online material comes to complement traditional paper material such as PowerPoint prints and book of exercises.

Several sections of this online documentation have been improved and continuously updated with particular focus on online exercises, and more explanations in conceptual parts. For instance, during the last semester of classes (February-June/2021), several parts of a section titled "How to use..." and related to instructions to execute simple exploratory statistical analysis using graphical calculators and R language (also used as a calculator), were incorporated and improved.

In the last class semester, and right at the beginning and for a total of 180 students, the online documentation was made available in a website (<http://sweet.ua.pt/pedrocruz/bioestatistica/> - in Portuguese). To evaluate the success of this initiative among students, an anonymous inquiry was carried out after final grades have been released. Due to the reduced number of responses (31 students), a semi-structure interview was also carried out a few students which were selected for claiming that they collected the document online. An important conclusion is that the online documentation should continue but also be improved.

Another goal was to keep a low time cost and effort to produce the online documentation. We have used Sphinx markup system that has proved its value in the community of Python programming language. In fact, a huge documentation of scientific computer libraries has been written online in this way. As for other markup systems, there is the possibility of producing documentation for printing with little extra effort.

Keywords: online documentation-based learning, statistics, Sphinx markup system

1 INTRODUCTION

The "World Wide Web" (Tim Berners-Lee, 1989 [1,2]) refers to interlinked documents by the so-called hyperlinks being the Wikipedia a well-known and well managed online encyclopedia system based on hyperlinks, where anyone can edit.

The Biostatistics course, at University of Aveiro, taught in Portuguese, started in 2010. Like most of university courses, study materials are the traditional PowerPoints for classes and a book of exercises. These materials start gaining form in 2010 when the subject started. These printed materials have evolved in first years and now course contents are stable with details that fulfil interests of most students.

Naturally, Wikipedia has entries with statistical contents, in Portuguese, but students usually want more focus on what is being under examination in their study materials.

The idea of creating "online documentation" for Biostatistics, with hyperlinks, appeared to capture all sorts of questions and answers from students to teachers. These interactions are mainly of the form "how do I do this?" and occur in classes or by email. Those questions are about exercises that were not always explicit given in class notes. Also, teachers receive questions of technical context like "How to do this in my calculator?" because using a graphical calculators or R system (for statistics) is compulsory. To help students to easily find answers, that are in the www "jungle", came the idea of an online documentation where students can find information in a structured way. By that time, a nice software (Sphinx [3], see technical section) called the attention of the first author by stating "Cheers for

a great tool that actually makes programmers want to write documentation!". With the start of the 19/20 pandemic the goal became stronger as the efforts to give students, at home, a complement to printed materials. Of course, there several other websites linked to universities courses, mainly in English language, and online education platforms but a dedicated online course will be preferably allowing focus on contents with Portuguese language.

The created online documentation is not (yet) a full "online learning course" [8] but a companion tool for students to autonomously find their answers and for teachers to easily give clues of what to read and were.

The target students come from Biology, Geology and Biology, and Marine Sciences degrees forming a group of 180 students distributed by two teachers (2020/2021). The course, in 2020/2021, was given mostly in live class.

2 METHODOLOGY

We follow a noninvasive methodological style: not heavily spread online documentation. A single hyperlink to the "online documentation" was made available in Biostatistics Moodle eLearning platform [4]. The platform is the landing webpage where students can find information of the course. The link points directly to the graphical calculators section of the "online documentation" that contains dedicated instructions to their use in the course. We thought that this information is the most sought. A second way to spread the "online documentation" was by email or MS Team chats. When a student asks something, a hyperlink is sent with information that the teacher think is useful. These two propagation methods were used in the second semester of 2020/2021 academic year. Communication by email, during the pandemic, was the preferable mean of communication outside the live classes.

We have used "Google Analytics" [5] to collect global usage statistics. In the Results section we provide a summary of number of accesses, length in pages and country percentage of users.

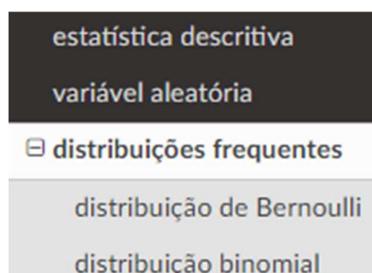
An inquiry, after last examination, was sent to students to acquire information of certain important aspects of the "online documentation". Since only 31 students have answered, a semi-structured interview was also carried with 3 students that gently answered our call.

To better understand the Google Analytics statistics, inquiring and semi-structured interview contents we first show what we have offered to students in the online documentation (freely available).

2.1 What are we offering to students?

As stated in abstract, one goal was offering a better study support and providing a means to share new explanations easily and quickly (via email, MS Teams or in class) when certain frequently asked questions arise. To do this, the "online documentation" has the following items described next.

A table of contents with pull-down menus with detailed subtopics. It can be seen main topics (in black background) and an open topic "*distribuições frequentes*" with open subtopics.



Visual appealing elements. Each page contains well drawn elements as they are necessary when mathematics is present: a nice mathematical notation, visual appealing tables and graphics, most of webpages have a subindex, lots of hyperlinks to easy the navigation between concepts. In the example below, we see a textual link in color, math notation from MathJAX [6], and a call for special attention "Dica" in green. The image at right side contains a nice rendering of R system instruction.

Assim, a função massa de probabilidade é

$$f(x) = P(X = x) = \begin{cases} 1 - p, & \text{se } X = 0 \\ p, & \text{se } X = 1 \end{cases}$$

Dica

A distribuição acima não é mencionada explicitamente nos slides.

```
pbinom(3, 20, 0.2)
[1] 0.4114489
```

A search box. Instead of walking in the table of contents, a search box tried to offer the user a direct link to the compiled information inside the website.



Contents about statistical concepts, with examples and procedures for R system and graphical calculator. The approach of the documentation is the same as a standard book, concepts with increasing levels of complexity with examples and also instructions for R software and graphical calculators.

distribuição binomial

Trata-se de uma **distribuição discreta** cuja descrição de contagem é sempre com esta forma:

X = número de «sucessos» em n provas independentes

Conteúdo

- [distribuição binomial](#)
 - [descrição](#)
 - [exemplos](#)
 - [propriedades](#)
 - [R project](#)
 - [Calculadoras gráficas](#)

Exercises with reading suggestions or solved exercises. The image shows a sub question and its reading suggestion and solution (sometimes containing the full answer).

(a) Qual o número médio de bactérias num cm^3 de água?

[sugestões](#) ▼

[solução](#) ▼

3 RESULTS

This section presents the “Google Analytics” basic statistical indicators, inquiry results and the important aspects from the three interviews.

During the semester, many emails were answered using replies with links to online documentation. In general, every reply was well acknowledged by students confirming one goals of online documentation: make easy to provide answers without having to repeat explanations.

By the time the second semester examinations were ending, in the starting of the summertime, an anonymous inquiry about online documentation and course framework were conducted (using google forms). Only 31 of 180 students of the course had answered. Because of that, a semi-structured interview was also carried out and only 3 students share their time with us.

The inquiry had 16 questions to be answered in 5-point Likert scale (numerical results and graphics were attached at the end). In the following paragraphs we interpret the results of the 31 answers and consider levels 4 and 5 as satisfactory. However, levels 1, 2 or 3 reveals that the situation must be analyzed in subsequent work.

3.1 How do you feel doing statistics?

To the question "Do you enjoy doing **descriptive statistical** treatment of data?" approximately 71% answered satisfactory and 90% valued doing the **statistical inference** (and not only descriptive treatment).

Considering that only a small fraction of all students answered the inquiry then these numbers suggest a possible problem of motivation in the course. Authors think that giving projects for statistical analysis is a next step to follow as it was well received by students in the past [7].

3.2 About the online documentation elements

Most students (93.5%) have consulted the online documentation in a useful manner, at least once. We will proceed with the analysis because online documentation could be useful.

Several aspects of the online documentation were presented in the last section and now we analyze results of some items:

- Table of contents: 93.1% of 31 students are satisfied. The organization of the table of contents and pull-down menus with subtopics is satisfactory.
- Search box: only 44.8% of 31 students are satisfied. Probably a better search technology could be necessary but since the table of contents is working this is not a priority.
- Quality of texts, concept explanations and procedures: 75.8% of 31 students are satisfied. This percentage value suggests the need to improve documentation.
- Satisfaction with graphical calculator instructions: 82,7% of 31 students are satisfied. It can be improved but is satisfactory because many students are using this resource. Since many of them study in groups then information will easily spread by students teaching their mates.

3.3 About exercises being online

During 19/20 and 20/21 almost all printed exercises were put online with sections to conduct the study. Two questions about this topic were:

- Study exercises could be only in online format: 82,7% of 31 students are satisfied. This value reinforces some usefulness of online documentation.
- The "online documentation" offers hyperlinks to required concepts to understand solution for each exercise. Students were questioned about two possibilities: these previous concepts (referenced by hyperlinks) should be before seeing full solution (51%) or in the middle of the full solution (49%). Since these percentages are well balanced our conclusion is that one should maintain both methods.
- The number of exercises is satisfactory: 89.7% of 31 students are satisfied. From our small sample size, that answered this survey, we could conclude the number is satisfactory but a more rigorous study with all students must be done considering that many could lack motivation by many factors.

3.4 About motivation to write online documentation

One student suggested that students could help improve documentation. The question was inspired by this idea but in this inquiry, possibly only that student answered satisfactory (level 4 or 5) (3.6% of 31). However, 32% answered level 3 opening this possibility but it's not a priority. We can consider putting this as an evaluated work like others that contribute to their final grade. In fact, in previous year few students contributed with calculated instructions.

3.5 Semi-structured interviews

Even the short number of interviews we gather useful information as presented next.

Question 1: “Did you used online documentation?” It is used as a complement to slides and exercises, before searching for teacher help. A student said also tried to find a proper book but he quit.

Question 2: “How did you use online documentation?” Class slides contains all concepts but sometimes they check for more details since slides could be more “abstract”. Online solved exercises were also consulted.

Question 3: “What’s your preference: only ‘apply rules’ or know more background of why to apply them?” Students did enjoy the application of rules without too much explanation because they get confused. However, a starting explanation is a good step to do.

Question 4: “Does the online documentation enhances cooperative learning?” No, online documentation is not enough.

Few other questions were vaguely answered but the from free talk several important aspects follow.

- Students don’t like to bother teachers in first place, they tend to ask help from mates; they used small groups in whatsapp.
- Many students were invited to apply Biostatistics inferential knowledge in another subject. The majority did not do that effort.
- A student mentions difficult in interpreting correctly the problem at hand.
- Students focus on their approval in the course and also to get a bigger classification, they are not so focused in learning theoretical aspects.
- Students use Excel in another related course.

3.6 Online documentation usage statistics

The following figures presents the number of different users accessing the online documentation over the second semester and also the mean minutes in the selected webpages. Values range from 80 different users to almost 190 different users (recall the universe of 180 students in the Biostatistics course). Overall mean minutes in the consulted webpages were 3:19 in the selected range. The red lines marks the five weeks of evaluation (five moments for each student with examination in his class period). Dates range from March 10 to March 24 of 2021. Weeks of examination were April 5 – 9, April 19 to 23, May 10 to 14, May 31 to June 4 and June 21 to 24 of 2021.



Figure Number of different users using online documentation by day (scale from 80 to 190 different users)



Figure Mean time in consulted webpages by day (scale from 0min to 10min)

Both figures suggest the students search for the online documentation, considering the peaks at examination days, and that they effectively use it, considering the mean time in webpages of online documentation. It was possible to see, from statistics from Aveiro city (were university is located), that the average time in pages was 11 minutes.

One can see a reduction in the number of users. The number of different students consulting the website decreased clearly before third examination. Several possibilities of explaining this drop out should be checked in future like economic reasons, lack of basic mathematical preparation and motivation.

Before second evaluation there is a peak. That is when they had to use calculators intensively in classes.

This online documentation is worldwide open so other Portuguese speaking countries can have benefit. As shown in next figure, Brazil presents the highest group of users after Portugal and Portuguese speaking countries only seldom visit the webpages.

1.	 Portugal	6 522		61,91%
2.	 Brazil	3 314		31,46%
3.	 Mozambique	356		3,38%
4.	 Angola	158		1,50%
5.	 Cape Verde	31		0,29%

4 CONCLUSIONS

Only a small fraction of enrolled students answered the inquiry presented in this work and, from these, only 70% of them say they enjoyed doing descriptive statistics. Authors consider these numbers as a possible warning of the problem of motivation in the course. Introducing group projects for application of statistics to datasets will probably help the motivation score for the course (not only tests every two weeks).

About the online documentation presented we collected information inclining to its usefulness, the table of contents is well organized and the graphical calculator instructions is presented in a satisfactory way. Study exercises in online format are welcomed by students so its notes about prerequisites with hyperlinks to concepts.

This research shows that more effort must be given to improve the quality of texts explaining concepts and driven by that more exercises considering helping more clear explanations about concepts.

During semester, more concern must be taken to prevent drop out in what is the teachers reach like lack of basic mathematical preparation and motivation.

ACKNOWLEDGEMENTS

Cruz and Freitas were supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020.

REFERENCES

- [1] Wikipedia contributors. "World Wide Web." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 11 Sep. 2021. Web. 14 Sep. 2021.
- [2] Wikipedia contributors. "Transmission Control Protocol." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 6 Sep. 2021. Web. 14 Sep. 2021.
- [3] Wikipedia contributors. "Sphinx (documentation generator)." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 23 Aug. 2021. Web. 14 Sep. 2021. <https://www.sphinx-doc.org/>
- [4] Biostatistics, moodle space: <https://elearning.ua.pt/course/view.php?id=322>
- [5] Google Analytics, <https://analytics.google.com/>
- [6] MathJAX, <https://www.mathjax.org/>

- [7] Amany K. Hassan, Stacie J. Lampkin, Timothy C. Hutcherson, "Students' perceptions of biostatistics following integration into an evidence-based medicine course series", *Currents in Pharmacy Teaching and Learning*, Volume 11, Issue 6, 2019
- [8] Castro, M.D.B., Tumibay, G.M. "A literature review: efficacy of online learning courses for higher education institution using meta-analysis," *Education and Information Technologies*, 26, pp1367–1385, 2021.