

# USING GA4 MEASUREMENTS TO CHARACTERIZE THE USE OF BIOSTATISTICS ONLINE SUPPORT DOCUMENTATION

J. Cruz, A. Freitas

*University of Aveiro (PORTUGAL)*

## Abstract

An Online Support Documentation has been built, since 2019, for a course of Biostatistics at University of Aveiro. It has been done to better support student study by complementing traditional slide presentations and a book of exercises. This course targets more than 180 students from the first year of the degrees of Biology, Geology and Biology, and Marine Sciences.

Last year, 28 students, out of 180, replied to an anonymous inquiry and a complementary interview was conducted with 5 chosen students. We could conclude that the development of the online support documentation should continue but improvements are needed (for instance, increasing the level of details in online texts).

To get a complementary view on how the Biostatistics online support documentation is used by students, a new approach is carried out. In 2020/2021 academic year we have automatically collected data usage using "Google Analytics 4" (GA4). The website, holding the support documentation, is a tree-like structure with chapters and sections. Each section is a webpage that can be of two types: a "concept" section or an individual "exercise" section. GA4 collects for each webpage: (1) the webpage title; (2) the number of interactions; (3) the mean time spent in all interactions with a webpage. An interaction (also called "User engagement") is issued by GA4 when a user has the webpage in focus for at least one second. Manually, two more columns were added to data: (4) the type of the webpage: concept or exercise; (5) the chapter number (1 to 6).

We present statistical insights about how the online support documentation was used in 2022/June when students had two formal tests about Chapters 4 and then Chapters 5 and 6. Briefly, our findings is summarized as follow: (1) as expected by the common sense, usage of webpages increased before assessments; (2) in Chapters 4 and 5, the "number of interactions", obtained for a webpage, and the "mean time spent", obtained for the same webpage, are positively correlated suggesting there are some pages more important in the learning process; (3) Concerning the mean time spent, its distribution is more skewed to the right for exercise webpages than for conceptual webpages. It could mean that the mean time spent on exercise webpages shows high variability because students look more for suggestions on the "how to do it" or to "confirm their solution" to a certain exercise.

Keywords: Online documentation-based learning, biostatistics, Google Analytics 4.

## 1 INTRODUCTION AND PEDAGOGICAL GOALS

To complement recommended books, slide presentations at classes and the book of exercises, we are developing, since 2019, an Online Support Documentation (OSD) [1] for our course of Biostatistics at University of Aveiro believing that easily accessible, dedicated written materials could better support the learning processes. Biostatistics subject belongs to the curricular plans of the first year of Biology, Geology and Biology, and Marine Sciences degrees. In general, more than 180 students are enrolled by academic year.

The composition of online support documentation satisfied the following structure:

- It is written in Portuguese because it is the students mother tongue. Erasmus students could easily get a "basic" translation using Chrome capabilities, for example.
- It presents a tree-like structure organized by "chapters". In each chapter, there are "sections". A section is a single webpage containing "topics". The following image shows a section called "Função de distribuição" and links for the related topics: "axiomas", "propriedades para a v.a. discreta", "propriedades para a v.a. contínua", and "notação".

**função de distribuição**

Genericamente, uma distribuição é uma função que atribui probabilidades.

A **função distribuição**, representada por  $F(x)$ , é o nome comum que se dá a funções que seguem a definição

$$F(x) = P(X \leq x)$$

sendo válida tanto para v.a. discretas como contínuas.

Por exemplo, num dado equilibrado a probabilidade de sair um número igual ou inferior a 4 é dada por:

$$F(4) = 1/6 + 1/6 + 1/6 + 1/6 = 4/6 = 2/3$$

**Conteúdo**

- função de distribuição
  - axiomas
  - propriedades para a v.a. discreta
  - propriedades para a v.a. contínua
  - notação

**axiomas**

Uma **função distribuição**, habitualmente representada por  $F(x)$ , deve obedecer ao seguinte:

Figure 1. The look of a section.

- There are two types of sections: “concepts” and “exercises”. The above webpage is, of course, a “concept” section. An exercise section (or webpage) as the layout, is exhibited in the next figure (where “sugestões” and “solução” means “suggestions” and “solution”, respectively).

**ex. 3.19**

A variância amostral corrigida do tempo de dissolução da droga no suco gástrico é de 2.69. Tendo em conta os seguintes quantis da distribuição  $\chi^2_{(7)}$ , obtidos com recurso ao software R

determine um intervalo a 95% de confiança para a variância do tempo de dissolução da droga no suco gástrico considerando que esse tempo é bem modelado por uma distribuição normal.

« sugestões ▼

« solução ▼

◀ Anterior

Seguinte ▶

Figure 2. The look of an exercise (“sugestões” are reading suggestions and “solução” is solution).

Our choice of technology to write our OSD is the markup language ReStructuredText (RST) [2] created for documenting Python language software. Since Python is being used in several fields of science, markup RST evolved to a system to communicate science, and this was a good reason to use it to write Biostatistics documentation. One important factor is the ability of the system behind RST (Sphinx library) to produce readable documentation for mobile phones, tablets, or desktop computers.

## 2 METHODOLOGY

Students were free to use the optional, not compulsory, OSD. This documentation was suggested to students in the beginning of semester, and, because it contains detailed instructions (in Portuguese and not in English) for four common models of handheld calculators (Texas and Casio).

Sometimes, when support was requested by students, using emails to the teacher, a link to a specific part of the documentation was given. The idea is common questions could be written and easily given to students under support request.

It is part of this “loosely methodology” to gather data from usage believing that if a student is using OSD, it is because he/she is learning a concept or a “how to do it” (or eventually checking).

The new platform “Google Analytics 4” was put to run, in each page of OSD, in middle of May/2022 so collected usage data is from that date.

## 3 RESULTS

In the academic year 2020/2021, 28 students, out of 180, voluntarily replied to an anonymous online inquiry. A complementary interview was also conducted with 5 chosen students. From both insights, we could conclude that the development of the OSD should continue but improvements were (are!) needed (for instance, increasing the level of details in online texts).

To get a complementary view on how the Biostatistics OSD was used by students, a new approach to gather data is carried out. In 2021/2022 academic year we have automatically collected data usage using “Google Analytics 4” (GA4) [3].

The website, holding the support documentation, is a tree-like structure with several “chapters” containing “sections”. A chapter is, like a book, an aggregation of sections and each section is a webpage (that can also be printed). Sections can be of two types:

- “Concept section” which contains a short summary, a table of contents for the topics related to the concept, an introduction of the concept, possible “how to” parts, related to mathematics notions (more theoretical notions of the section);
- Individual “exercise section” which corresponds to a single webpage, containing an exercise and its items (a), (b), etc, a “pop up” with suggestions for readings and another “pop up” with solutions. More advanced exercises have also a method to solve the problem.

As stated, Google Analytics 4 “is an analytics service that enables you to measure traffic and engagement across your websites and apps” [3]. The “engagement” is an interaction of the user with a webpage (also called “User engagement”) and is issued, by GA4, when a user has the webpage in focus for at least one second. Overall, GA4 collects, for each visited webpage, the webpage title, all interactions, and the duration of each in seconds.

Because of a limitation of the availability of GA4, we gathered data only between may/22 and june/25 (2022). During this period, students did two random computerized formal tests (out of a total of five) that most students were called to do. Those examinations were about Chapter 4 and then, Chapters 5 and 6.

After downloading of the above data, we additionally added information about each visited webpage related to the Biostatistics domain: the “type” of the webpage: “concept” or “exercise”, and also, the chapter number that we ranked from Chapter 1 to Chapter 6.

The data file has the following columns:

1. Webpage title: it is the section name like “Binomial Distribution” or “Exercise 1.12”.
2. Total number of interactions (G4 measured) in the period of study.
3. Average time, over the period of study, spent on each webpage (considering only those spending more than 1 second).

4. Type of webpage: concept or exercise.
5. Chapter number (1 to 6).

Statistical insights about how the OSD was used in 2022/June when students had two formal tests about Chapter 4 and then Chapters 5 and 6.

A summary of this findings is as follows:

**Finding 1.** As expected, by common sense, usage of OSD for Biostatistics increased before assessments: Chapter 4 at may/31 and Chapters 5 and 6 later with students divided in two groups: one at June/21 and other at June/23. The following plot presents the mean time by user in each day.



Figure 3. Overall usage mean time, by day.

**Finding 2.** For each webpage, concept or exercise types, the "number of interactions" and the "mean time spent" are positively correlated suggesting an important ranking for each webpage. This conclusion visually arises from the scatterplots in the next figure where each webpage is characterized by a pair (number of interactions, mean time spent). We conclude that the teaching and learning process needs to pay attention to higher values of (number of interactions, mean time spent). Considering only the chapters under evaluation, the positive correlation is significant for the "complex" Chapters 4 and 5 but not for "simpler" Chapter 6 (scatter graph not shown).

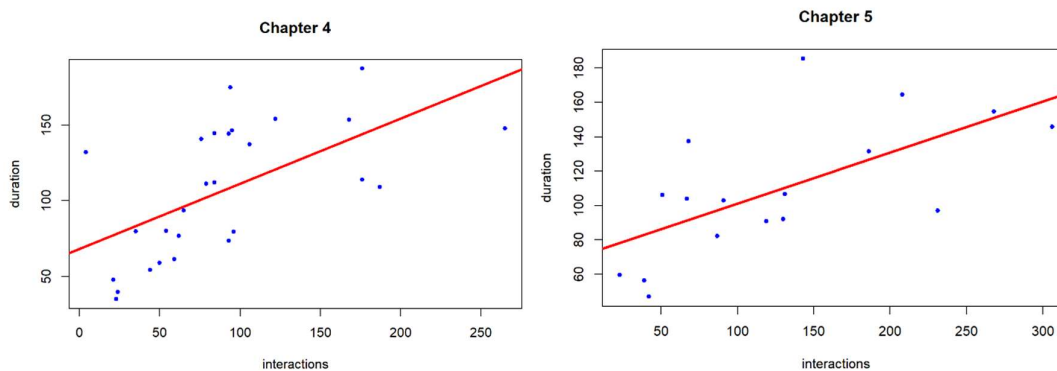


Figure 4. Webpage interactions versus mean duration in each webpage

**Finding 3.** Concerning the mean time spent, its distribution is more skewed to the right for exercise webpages than for conceptual webpages. It could suggest that the mean time spent on exercise webpages shows high variability because students look more for suggestions on the "how to do it" or to "confirm their solution" to a certain exercise.

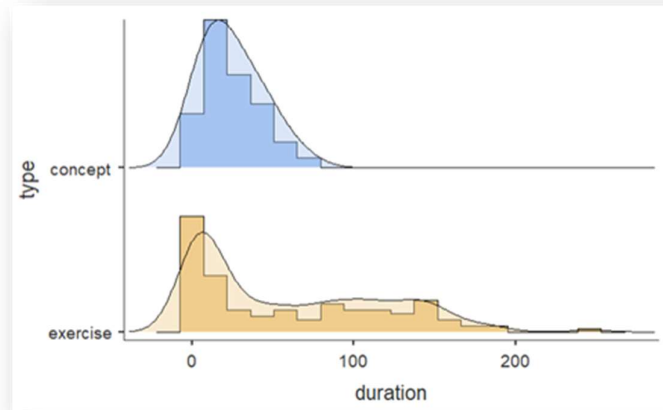


Figure 5. Histograms of duration versus concept or exercise webpages

#### 4 CONCLUSIONS AND FUTURE WORK

This work presents a second evaluation of the OSD, this time using online records of interactions and time spend on webpages. The first evaluation used inquiries and interviews. Both evaluations suggest keeping this OSD alive and improve it. Three main findings: (1) usage of webpages increased before assessments; (2) more mean time spent in some webpages suggests they are more important in the learning process; (3) mean time spent for exercise webpages tends to be higher than for conceptual webpages suggesting more time in "how to do it" or confirm some exercise solution.

Engagement is measured by interactions. Then, the next step is to include, for each webpage, a "Like" / "Don't like" buttons as a second measure of engagement. Webpage with concepts and exercises with higher values of (number of interactions, mean time spent) maybe need more careful attention in presentation slides and maybe more simpler exercises in order to ease the absorption of matters. This method should be repeated in coming year to evaluate other chapters.

#### ACKNOWLEDGEMENTS

This work is 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), reference UIDB/04106/2020.

#### REFERENCES

- [1] Online Support Documentation (OSD) for a Biostatistics course. <http://sweet.ua.pt/pedrocruz/bioestatistica/>
- [2] ReStructured (RST) and Sphinx Python. <https://www.sphinx-doc.org/en/master/>
- [3] Google Analytics 4. <https://support.google.com/analytics/answer/11109416#user-engagement>