# INTERDISCIPLINARY AND DIFFERENTIATED LEARNING OF STATISTICS IN AN ONLINE MPH

Marta Avalos Fernandez and Gaëlle Coureau
SISTM team INRIA BSO, Talence, France
University of Bordeaux, BPH Research Center, Bordeaux, France
marta.avalos-fernandez@u-bordeaux.fr

*Statistical literacy is necessary for evidence-based public health practice. However, introductory statistics courses can be challenging and stressful for continuing education students. The Master of Public Health program at the University of Bordeaux offers an international open and distance learning program to address the increasing demand for public health education from working professionals who have limited time for training or geographical constraints.*
*To enhance the learning experience, the program has developed problem-based approaches based on epidemiological studies and data by integrating statistics and epidemiology courses. Additionally, the program adapts coding learning objectives to students' skills and career plans by teaching R with two interfaces. While RStudio is recommended for students planning to enter data sciences as it requires more advanced coding skills, R Commander is suitable for students with weak computer skills and those who don't plan to evolve in data sciences. This strategy has proven to be an effective compromise.*

## CONTEXT

The first year of the Master of Public Health (MPH) program at the University of Bordeaux (France) has been organized in Open and Distance Learning (ODL) since 2007. The second year of the online MPH program was launched in 2016. Courses are aimed exclusively at French-speaking continuing education students, mostly, professionals working in the health field or in the social and solidarity economy sector. In 15 years, more than 300 professionals have been trained, about half of whom come from Africa and the other half from overseas departments and regions of France, French-speaking countries in Europe and abroad. The program aims to meet a growing demand in public health from active professionals whose mobility and/or time for training are limited.

Continuing education students are adults who return to learning after leaving initial education and training to enhance their employment prospects, to develop professionally and to obtain transferrable skills. In France, continuing education has been part of the missions of universities since 1968. Its legislative framework has evolved since then, the latest reform dates from 2018, and is intended to boost continuing education by allowing everyone to better fit into their career plans as well as to enhance equity between men and women, and full-time and part-time employees. More generally, in the European Union, the need to significantly increase adult learning has been pointed out which has leaded to the emergence of the concept of "lifelong learning". However, access to training remains difficult since it requires a significant investment on several levels (particularly, financial and organizational) which often leads to inequalities. To face unequal access, surveys recommend an adapted pedagogical model, in terms of content, rhythm of adults (encouraging distance learning, part-time work), digital uses, competencies, etc (Santa Maria da Feira European Council, 2000).

The African context, including public health issues, is completely different. In 2007, Africa had 1.3% of the world's trained public health personnel while it had 25% of the disease burden (Mokwena, 2007). On the other hand, 55% of African countries did not offer postgraduate training in public health (Ijsselmuiden et al., 2007). With the development of the Internet (and subsequently, online teaching programs), distance learning via the internet by Western institutions appeared as an alternative to the shortage of training (Alexander et al., 2009). A partnership has been established with the University Agency for French-speaking communities (AUF) which partially finances several students per year. It should be noted that over the last decade, many African universities have built and improved training programs that meet local needs, including online programs (Amde et al., 2014).

Statistical literacy and knowledge of statistical concepts and methods are necessary for understanding the public health literature, a key step in evidence-based public health practice. Therefore, critical evaluation of the application, presentation, and interpretation of statistical analyses

in public health studies is an essential skill for MPH degrees (Calhoun et. al., 2008 or Oster & Enders, 2018). Yet, introductory statistics courses require strong skills in scientific reasoning, abstract thinking, mathematical calculation, computer manipulation, etc. These educational requirements have the potential to induce anxiety related to mathematics or programming skills (Luttenberger et al., 2018, Forrester et al., 2022, Miller & Pyper, 2023). They are particularly stressful and a source of failure for continuing education students who have long since left the university and have started again in ODL.

Public health is a field of choice for developing interdisciplinarity. Indeed, targeting population health improvement involves a variety of complementary disciplinary approaches. To enhance the learning experience of introductory statistics, the program has developed problem-based approaches based on epidemiological studies and data by integrating statistics and epidemiology courses.

International classes are made up of diverse student profiles with heterogeneous backgrounds and prior knowledge. In addition, during the second year of the MPH, students choose a specialization that will provide them room for career advancement in their chosen area of interest. Two options are offered: health promotion and data management and analysis. The first option trains students for designing public health interventions to improve populations health. The second option is designed to train students for utilizing quantitative methods for collecting, storing, retrieving, analyzing, and interpreting health data in conducting epidemiological studies. This option requires more advanced coding skills. Coding learning has been differentiated to meet the needs, preferences and goals of individual students.

## INTEGRATION OF STATISTICS AND EPIDEMIOLOGY COURSES: UNITY IS STRENGTH

The teaching period runs from September to December (first semester) and from January to March (second semester). Students then begin their internships, with a minimum duration of 2 months for first-year MPH students and 4 months for second-year MPH students, respectively. The course schedule including the online availability of documents on Moodle, theoretical courses, exercise corrections, assignment corrections and deadlines for graded and non-graded assignments, past years' exams, synchronous learning sessions (via the video conferencing tool zoom) etc., is announced at the beginning of the semester.

Since 2016, there has been a gradual bringing together of statistics and epidemiology courses in order to develop problem-based approaches based on epidemiological studies and data. Statistics are thus presented as an essential tool for managing uncertainty in epidemiology and, on the other hand, the context of epidemiology makes it possible to give meaning to the application of statistics, thus improving understanding. The courses are co-leaded by a biostatistician and an epidemiologist. At the beginning, the integration of statistics and epidemiology courses consisted only in sharing the contents coherently. At present, the organization of the statistics and epidemiology courses is the following:

- First year, first semester (60 hours). Statistical courses contain: descriptive statistics, general terminology and basic concepts in probability and statistics, estimation and confidence intervals, general principles of statistical hypothesis testing

One of the graded assignments and the final exam are commonly redacted by the two teachers. In addition, both teachers animate together the video conferences preceding them.

- First year, second semester (90 hours). Methods in epidemiology, statistical tests and regression methods. Statistical courses contain: parametric and non-parametric statistical tests, linear and logistic regression methods

One of the graded assignments are commonly redacted by the two teachers.

- Second year, first semester (60 hours). Advanced methods in etiological epidemiology. Statistical courses contain: logistic, conditional logistic and Poisson regression methods

The graded assignment and the final exam are commonly redacted by the two teachers. In addition, both teachers animate together the course of support to the graded assignment.

- Second year, second semester (60 hours). Epidemiological surveys and data analyses from cohort studies. Statistical courses contain: methods for survival data analyses. These courses addressed to students who plan to enter data science, they are taught independently.

## ADAPTING CODING LEARNING OBJECTIVES: DIVIDE AND RULE

Initially, between 2007 and 2011, the programs taught for database management and statistical analysis were EpiInfo, a user-friendly software widely used by epidemiologists, and MS Excel spreadsheets. Occasionally, students with sufficient computer background received a CD containing a SAS student's license to explore the software on their own. In 2011, SAS was replaced with R (R Core Team, 2023), while still offering the option to work with EpiInfo. Starting from 2013, Rmarkdown was utilized to create HTML and PDF tutorials. The progress was gradual but slow, as learning the reasoning and statistical methods was already stressful and challenging for our students, and incorporating programming in R introduced additional anxiety.

In 2017, the University of Bordeaux developed an online tutorial called Begin'R to address the difficulties faced by on-campus students when learning R programming with the RStudio interface. Additionally, a tutorial for R Commander – Rcmdr – (Fox & Bouchet-Valat, 2022), tailored to our course content, was created in French. Although a video conferencing platform had been in use since 2014, it was during the COVID-19 pandemic that significant improvements were observed in the available platforms. Adopting the Zoom platform allowed for screen sharing by students, and even the teacher could take control of their computer if necessary, resulting in a tangible enhancement in the distance learning of statistical software.

RStudio requires more advanced coding skills compared to Rcmdr (Wilson, 2012, Abbasnasab et al., 2021). Hence, it is recommended that students with limited computer skills begin with Rcmdr and gradually transition to RStudio if they plan to pursue data science. Not all students have their career plans completely defined when they start the Master's program. Generally, we recommend the following progression:

- In the 1st semester of the 1st year, an initial introduction to Rcmdr ensures that all students have correctly installed the software, can load files of different formats, perform variable transformations and sample selections, and apply descriptive statistics. At this stage, no evaluation of software mastery is conducted.
- In the 2nd semester, all statistical methods such as hypothesis testing, linear regression, and logistic regression are taught using Rcmdr. During videoconference courses, all students, along with the teacher, perform exercises using Rcmdr. Students who already possess some knowledge of RStudio can use it for the manipulations. Whether working with RStudio or Rcmdr, if students encounter any problems, they can share their screen, and the teacher can provide guidance. At the end of the course, the report generated automatically with Rcmdr from Rmarkdown is made available to the students. This report is useful for both RStudio and Rcmdr users. Additionally, a brief video is recorded summarizing the main steps, pitfalls, errors, or misunderstandings identified during the session with the students.
- During the 1st semester of the 2nd year, separate videoconference courses are conducted for teaching RStudio and Rcmdr. One month after the start of the course, an ungraded placement assessment is administered. If they prefer it, non-data science-oriented students will solely employ the Rcmdr interface. Students who already have a sufficient mastery of RStudio and aspire to pursue data science will use RStudio. Lastly, students who intend to enter data science but have limited computer skills are encouraged to work with both interfaces and rely on the Rcmdr script to facilitate their transition to RStudio. The objective for these students is to become comfortable with RStudio by the final semester and for their internship work.
- During the 2nd semester of the 2nd year, only students who have selected the data science option proceed with their biostatistics and epidemiology courses. These courses are exclusively conducted using RStudio.

In addition, switching between the two interfaces is facilitated by utilizing the same packages, for example, gtsummary and tidyverse, for constructing descriptive tables and effectively presenting regression results.

## CONCLUSION

In addition to direct interactions with students throughout the year, each semester an anonymous satisfaction survey is sent to students. Analyzed together with the results of course assessments, student’s success/failure, student’s internship work (for data science students), the survey constitutes a valuable source for the evolution of practices. All these feedbacks have led to the

increasingly close integration of courses and differentiating learning. Here are a few examples of modifications made in response to surveys or poor results:

- Initially, we believed that presenting epidemiology and statistics courses together would be sufficient for students to establish the connection between the two disciplines. However, this approach did not effectively facilitate the desired link. Therefore, we restructured the courses, integrating them more intricately, and sharing sessions, to explicitly highlight the connection.
- We assumed that tutorials (such as Begin'R or addressing Rcmdr use) would adequately equip students to independently master the software. This assumption was based on the experience of in-person courses, where students benefit from collective dynamics. However, we realized that these resources were insufficient for distance learning students who lack the same peer interaction. Consequently, we recognized the need for closer support to promptly address any obstacles students encountered.
- Even when students appeared to follow the software courses adequately, we found that a significant number of live sessions were essential for achieving basic mastery. This approach ensured that students did not remain stuck for prolonged periods, as immediate assistance during live sessions facilitated quicker problem-solving.

These strategies appear to be presently an effective compromise. Nevertheless, this sector requires constant re-evaluation, re-adaptation and re-thinking of the best educational strategies and learning tools. Thus, new ChatGPT coding assistants for RStudio or RStudio evolution to integration of R and python languages could completely change the panorama.

REFERENCES


Abbasnasab Sardareh, S., Brown, G.T.L. & Denny, P. (2021). Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences. *Teaching Statistics 43*, S157– S172.

Amde, W.K., Sanders, D. & Lehmann, U. (2014). Building Capacity to Develop an African Teaching Platform on Health Workforce Development: a Collaborative Initiative of Universities from four sub Saharan Countries. *Hum Resour Health. 12*(1), 31.

Alexander, L., Igumbor, E.U. & Sanders D. (2009). Building Capacity without Disrupting Health Services: Public Health Education for Africa through Distance Learning. *Hum Resour Health. 7*(1), 28.

Calhoun, J.G., Ramiah, K., Weist, E.M. & Shortell S.M. (2008). Development of a Core Competency Model for the Master of Public Health Degree. *Am J Public Health. 98*(9), 1598-607.

Forrester, C., Schwikert, S., Foster, J., & Corwin, L. (2022). Undergraduate R Programming Anxiety in Ecology: Persistent Gender Gaps and Coping Strategies. *CBE life sciences education*, 21(2).

Fox, J. & Bouchet-Valat, M. (2022). Rcmdr: R Commander. R package version 2.8-0, https://socialsciences.mcmaster.ca/jfox/Misc/Rcmdr/.

Ijsselmuiden, C.B., Nchinda, T.C., Duale, S., Tumwesigye, N.M. & Serwadda, D. (2007). Mapping Africa's Advances Public Health Education Capacity: the AfriHealth Project. *Bull World Health Organ. 85*(12), 914-22.

Luttenberger, S., Wimmer, S. & Paechter, M. (2018). Spotlight on math anxiety. *Psychol Res Behav Manag*. 11:311-322.

Miller, A. & Pyper, K. (2023). Anxiety Around Learning R in First Year Undergraduate Students: Mathematics versus Biomedical Sciences Students. *Journal of Statistics and Data Science Education.* DOI: 10.1080/26939169.2023.2190010

Mokwena, K. (2007). Training of Public Health Workforce at the National School of Public Health: Meeting Africa's Needs. *Bull World Health Organ. 85*(12), 949-54.

Oster, R.A. & Enders, FT. (2018). The Importance of Statistical Competencies for Medical Research Learners. *J Stat Educ. 26*(2), 137-142.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Santa Maria da Feira European Council 19-20 June 2000: Conclusions of the Presidency [Internet]. [cited 30 April 2023]. https://www.europarl.europa.eu/summits/fei1_en.htm#II

Wilson, J. (2012). Statistical computing with R: selecting the right tool for the job—R Commander or something else? *WIREs Comp Stat. 4*, 518-526.