

OPEN EDUCATIONAL RESOURCES FOR STATISTICS TEACHING

James Musyoka¹, Roger Stern² and David Stern³

¹Maseno University, Department of Statistics and Actuarial Science, Kenya

²Statistics for Sustainable Development, UK

³Statistical Services Centre

jkmusyoka@maseno.ac.ke

A series of papers at this conference have built on the African Data Initiative (ADI) which was started to improve statistical literacy and understanding. One immediate deliverables of this initiative has been new statistical software which is free, easy-to-use, open source and which encourages good statistical practice. Here we show how this software together with other open educational resources can be used to improve statistics teaching. This is demonstrated using an undergraduate course which was offered to about 300 students at Maseno University, Kenya. The resources include those from an e-learning course, called e-SMS (Statistics Made Simple) and an electronic statistics book called Computer Assisted Statistics Textbook (CAST). The course also made extensive use of Moodle to enable a “blended” approach to be undertaken.

INTRODUCTION

In Africa, statistics is still taught using the theoretical approach which puts more emphasis on the mathematical formulae rather than the concepts. This claim was recently corroborated by a set of 54 top students in Africa admitted to undertake a Master of Science degree in Mathematics Sciences at one of the African Institute of Mathematical Sciences (AIMS) centers based in Tanzania. Through a short survey, most students (32) revealed they did not use any statistical software. The largest dataset used by any of the respondents had just 50 rows (cases) (Fundu, Clements, Stern, Stern, Renaud, & Sananka, 2017). It was therefore not surprising that the students had little or no data handling skills.

Within Kenya there has been some awareness that this needs to change for many years (Odhiambo, 2002) and internationally the drive for change is gathering momentum (Cobb, 2015) with guidelines supporting a less theoretical approach (Carver, et al., 2016). Some of the challenges that contribute to the continued theoretical approach in statistics teaching in Kenya include lack of availability of statistical software that is suitable for teaching and also lack of teaching materials that build statistical understanding. Without software that is free and easy-to-use institutional adoption is highly problematic. Given the large class sizes, it is also vital that materials to teach of statistics courses less theoretically are interactive to allow the students to practice and check that they have acquired the correct understanding of the concepts, without requiring large amount of personal lecturer interaction.

Because of these challenges the African Data Initiative was started, with the first activity being to develop the statistics software needed to fill the existing gap (Stern, 2017). The software will be officially launched at the 61st World Statistics Congress at Marrakech, Morocco. In addition, teaching materials previously used in an online statistics course which has been effective in improving statistical understanding among both statisticians and non-statisticians (Dale, Clark, Stern, Leidi, & Stern, 2010), were recently released as open educational materials, (SSC, 2015). These materials can now be used in their original format or adapted by lecturers in their statistics teaching.

In this paper, we describe both of these resources. Through an undergraduate statistics course which used these resources, we also demonstrate their potential to improve the teaching of statistics. A blended approach to teaching was adopted to make fuller use of the resources. An electronic statistics textbook called CAST (Computer Assisted Statistics Textbook) which has been available to help transform the teaching of statistics, (Stirling, 2010) was also used as an additional reading resource. This paper aims to raise awareness of these open educational resources tempting other statistics educators to use them in their teaching, sharing their experiences and improvements.

THE UNDERGRADUATE COURSE

The undergraduate statistics course that used these resources is called ‘Statistical Computing II’. It is one of the four skill-based courses offered to third year students in the School of Mathematics, Statistics and Actuarial Science at Maseno University. The courses are quite new in the curriculum and were added as part of an initiative to enrich the statistics content with practical skills, (Stern, Ongati, Agure, & Ogange, 2010). The previous curriculum comprised theoretical courses only which did little to build data skills and statistical understanding among the students.

“This course reviews the comparison of a spreadsheet with statistical packages, (introduced in Statistical Computing I) and introduces a range of statistical packages available at the University. The value of being able to use commands in a statistical package as well as dialogues is described, and tasks from Statistical Computing I are reviewed through commands. Data are then generated from examples of problems, such as a Markov chain, or a randomized block design. The resulting data are then analyzed using the statistical software. Multiple samples are then used to illustrate concepts such as confidence limits and significance tests. The ideas are compared with the equivalent sections in an electronic textbook, such as CAST. The importance of randomization is explained. The statistical software is used to generate randomized designs for a range of designed experiments, and for simple and stratified surveys”

Figure 1: Course Description.

As shown by Figure 1, the course description includes a range of statistics topics most of which are usually covered before the students get to third year. The aim of the course is therefore to give the students some practical experience to these topics as previous courses were mainly theoretical. Hence the course needed a new set of materials and resources that can achieve this objective.

TEACHING MATERIALS

One of the new set of materials used in the teaching of the undergraduate course was a set of open educational materials developed from a successful online statistics course. The online course is called Statistics Made Simple and was originally designed for non-statisticians (first-year postgraduates and professional) at the University of Reading whose roles and/or studies required core skills for statistical analysis. But the materials were later found to be a useful resource to statistics students at MSc level from Africa whose teaching did not help very much with the understanding of statistics concepts. The lecturer for the undergraduate course is one of the many beneficiaries of the course. This was possible through an existing collaboration between Maseno University and the University of Reading, (Musyoka, Stern, & Stern, 2014)

The materials have been effective for statistics training to both statisticians and non-statisticians largely because they are not formulae-oriented. They instead focus on helping students develop skills for translating a real-world into a statistical problem, understanding the appropriate statistical methods for solving the statistical problem, interpreting the results of a statistical analysis and relating the solution back to the real-world problem. This approach is consistent with that advocated by (Wolfram, 2010) as an effective way to teaching maths.

The materials are packaged in interactive presentations that are easy to download and interesting to read and follow. Plausible characters are used in each presentation to mimic a class discussion which was useful in encouraging active learning during the lecture sessions for the undergraduate course. Some presentations also contain short questions which learners are expected answer to check their understanding before continuing with the presentation. Learners are also able to pause and go back to review the materials as they wished.

The statistics topics for which the presentations exist are shown by Figure 2. The presentations are available on the Statistical Services Centre website and can be accessed using this link <http://www.reading.ac.uk/ssc/resourcepage/materials.php#default>. The presentations can be viewed online, but there also exists offline versions which can be downloaded for offline use. The

offline versions of the presentation are editable hence can be adapted for any course whose objectives can be met by these materials.

- **Topic 1:** Introduction - What is Statistics?
- **Topic 2:** Describing Data Well
- **Topic 3:** Making Good Generalisations
- **Topic 4:** Hypothesis Testing and P-Values
- **Topic 5:** Introduction to Statistical Models

- **Topic 6:** Explaining Variability
- **Topic 7:** More on Explaining Variability
- **Topic 8:** Non-Normal Data
- **Topic 9:** Generalising Modelling Ideas
- **Topic 10:** Coming Down from the Mountains

Figure 2: Statistics topics.

About this CAST e-book	
0. Preface	
1. What is Statistics?	
2. Describing Data Well	
3. Making Good Generalisations	
4. Testing Hypotheses	
5. Statistical Models	
6. Categorical Variables	
7. Explaining Variability	This book is a
8. Non-Normal Data	concepts that y
9. General Modelling Ideas	
10. Coming Down from the Mountains	This CAST bo
Index	So, the e-sms i
Datasets	each week will

Figure 3: CAST e-book contents.

CAST BOOK FOR SMS

To supplement the open educational resources, an electronic statistics resource called Computer Assisted Statistics Textbook (CAST) was also used in the undergraduate course. CAST is a collection of electronic books that mainly cover introductory statistics topics using data and scenarios from different application areas. The books also contain exercises and tests which together have been used to successfully improve statistics teaching at the tertiary level in Kenya, (Manyalla, Mbasu, Stern, & Stern, 2014).

The specific CAST book that was used in this initiative is called CAST eBook for eSMS. This book was designed specifically for participants of the online course. As shown by Figure 3, the contents of book are categorised into the same ten topics of the online course. The book can be found under the custom eBooks tab of the CAST homepage. This is the holding place for all customized versions of the CAST book. While it is possible to develop a new CAST book for a given statistics course, it is much easier to re-use materials in existing books.

CAST was instrumental in simplifying statistics concepts making it easy for students especially those who were not statistics majors to understand. This was possible through the dynamic graphics and interactive materials in the book which allowed the students to vary examples in an effort to master the concepts. Since CAST uses datasets in its teaching, the experience was therefore practical and hands-on for the students.

The experience was however not without challenges. Downloading the book for initial installation was challenging for most students because of limited internet connectivity at the University. The installation process was thought to be lengthy as it also involved the installation of Java which is needed to run the dynamic graphics. Hence, the amount of technical support required was initially a substantial responsibility to the lecturer who doubled up as the technical support person. But there were solutions to these challenges that could have made the experience much smoother if they had been thought ahead of the course. One such solution was to supply the students with both the CAST and Java installers in a compact disc. This helped with the internet challenges that had threatened to make the use of the electronic book impossible.

STATISTICAL SOFTWARE

The new software developed by the African Data Initiative, R-Instat, was used for the first time in the undergraduate course. The role of the software in the course was to give the students a hands-on experience of analysing real data. This is something they rarely experience in other courses. Practical assignments developed for the online course were adapted to support the use of R-Instat.

The main software window, shown in Figure 4, has two sub-windows. The data view sub-window has the capability of holding various datasets in different sheets. This feature was used in

the early stages of the course to illustrate that in R-Instat one can handle more than one dataset at the same time.

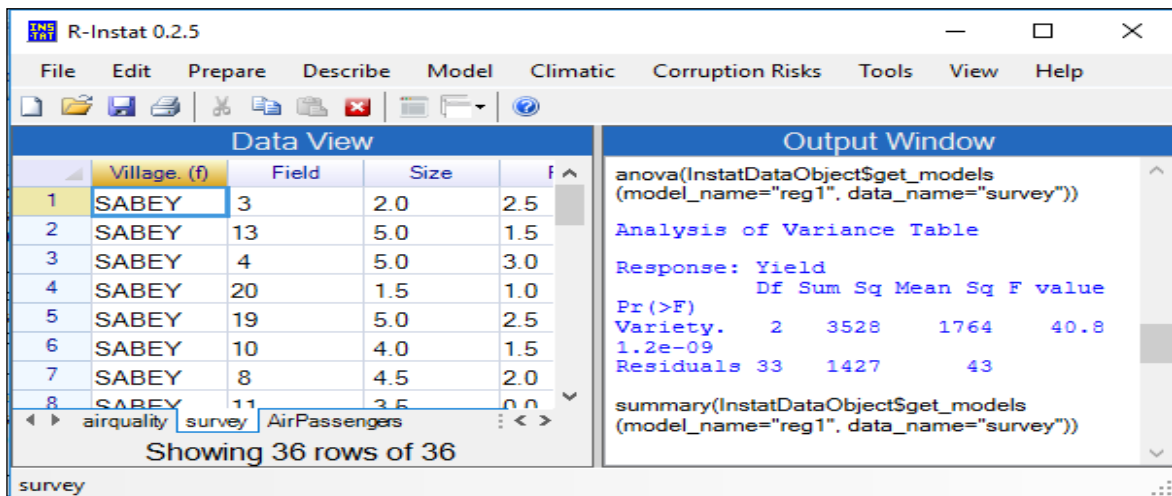


Figure 4: Main software window.

There are three menus in R-Instat which were relevant for the statistics topics covered by undergraduate statistics course. The Prepare menu was used to teach principles and practice of data manipulation in preparation for analysis. This idea built on a pre-requisite course about data handling using spreadsheets. Detailed information about the prepare menu and how it can be used in statistics teaching are provided by (Lunalo, et al., 2017).

The other two main menus were Describe and Model. The Describe menu was used to summarize columns of data using tables and graphs. (Fundu, et al., 2017) explain the importance of this menu to support statistics teaching. The Model menu was used to perform regression analysis and produce analysis of variance tables. This was required for topics 5 – 9 of the course.

THE BLENDED APPROACH

To make full use of the resources described above, the course was delivered using a combination of face-to-face sessions and online learning. This was the first time that a blended approach was used to teach a regular course at the University.

The online learning involved using a learning management system Moodle to deliver the pre-designed course materials to the students. The system was also used for management and grading of assignments through quizzes. Moodle also provided forums which facilitated interactions between the students themselves and also with the lecturer.

The face-to-face sessions which lasted 3 hours every week were mainly used for discussions about the materials for the week and practical sessions using R-Instat. The discussions were only possible because the students had agreed to study the materials online prior to the lectures. The discussions therefore were designed to improve their understanding.

This flipped classroom approach was new for both the lecturer and students. It is notable for the size of the class, which consisted of around 300 students. Another lecturer at Maseno had previously flipped their class in this way with similar resources but without the benefits offered by the Moodle course management system (De Giusti, 2017).

STUDENT EVALUATION

At the end of the course, an evaluation questionnaire was administered to obtain student's opinion about the course, the resources, the teaching approach and the entire experience. The questionnaire was voluntarily administered through the Moodle system. Out of 300 students, 163 participated in the survey. Non-response was considerable partly due to internet connectivity issues during the last week of the course.

The first question asked the students to describe the entire experience and a majority (155 of the 163, 95%) thought that the course was interesting, with most of them also finding it challenging. See the rest of the results as shown by Figure 5.

Responses	Interesting and easy	Interesting but with challenges	Nothing out of ordinary	Not interesting but easy	Not interesting and with challenges	Total
	25 (16%)	127 (79%)	5 (3%)	2 (1%)	1 (1%)	160

Figure 5: Survey Results.

The students' opinion about how useful the open educational materials were showed that a majority (around 92%) generally found these materials were useful in understanding statistical concepts. However out of these, only 25 (16%) said the presentations were the most useful resource in the course which is reassuring because they are the hardest to produce.

A majority of the students also found the practical assignment helped to understand the ideas introduced by the open educational materials. On a scale of 1=not at all up to 5=a lot, the average rank was over 4!

Students' opinions about the four characteristics of R-Instat were also sought. Most of the students agreed that the software is designed to help understand some statistics principles and also that the software is powerful. However a considerable number of students (40%) neither agreed nor disagreed that the software was easy to use. Figure 6 presents the rest of the results.

Responses	Strongly Disagree	Disagree	Average	Agree	Strongly Agree	Total
Easy to use	6 (4%)	12 (8%)	63 (40%)	62 (39%)	16 (10%)	159
Designed to help understand some statistical principles	3 (2%)	3 (2%)	29 (18%)	77 (48%)	47 (30%)	159
Powerful	6 (4%)	7 (4%)	36 (23%)	80 (50%)	30 (19%)	159
Easy to install	13 (8%)	15 (9%)	49 (31%)	53 (33%)	29 (18%)	159

Figure 6: Survey results about R-Instat.

The students were also asked whether they recommend that the blended approach (A mixture of face-to-face and online) should be used in other courses at Maseno University; a majority (82%) supported the idea.

The last question asked for comments about the course. Most of the comments were positive, but a few were negative particularly because of restricted access to IT infrastructure. The most interesting brought to our attention delivery issues related to the blended method being abused by some students who did not work on the weekly assignments themselves. However it was noted that this left them unprepared for the exam. It will be interesting to see the end results after marking the exam.

CONCLUSION

In this paper, we have described some of the open educational resources that can facilitate the much-needed transformation of statistics teaching in African Universities. The resources include a new statistical package, open teaching materials and an electronic statistics book all of which are freely available for use by both statistical educators and learners. The resources cover a wide range of topics in statistics hence can be used in many courses. Some of the resources can be adapted to suit the objectives of a specific course.

Through our experience, we found that teaching in this flipped way using these resources was not more time consuming than teaching in more traditional ways. This is essential because of the heavy teaching loads that each lecturer has to deal with every semester. The resources provide an opportunity for lecturers to shift their time and effort which enabled more of interaction to focus on building understanding of the concepts amongst learners.

This course did involve more effort and work by the students, compared to their other modules. There were particular challenges because all the innovations (software, learning management system, electronic textbook and e-learning materials) were introduced in a single course, so the students had to master both the technologies as well as the subject. Hence we were surprised at the relatively high proportion of students who endorsed the approach. Their comments and discussions confirmed that they welcomed the improved skill set, and could see its relevance for their future work.

This course and the way it has been delivered is part of a continued effort to improve statistics teaching at Maseno University. It is particularly notable because it is the first time that teaching innovatively has not been more demanding than the traditional approach. This is a substantial advance towards building innovations that can scale virally (Stern D. , 2014). The next big challenge is get others to adopt these innovations in their teaching so that the data revolution (UN, 2014) can really take off.

REFERENCES

- Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., et al. (2016). Guidelines for Assessment and Instruction in Statistics Education (GAISE). *American Statistical Association*.
- Cobb, G. (2015). Mere Renovation is Too Little Too Late: We Need to Rethink our Undergraduate Curriculum from the Ground Up. *The American Statistician* , 69 (4), 266-282.
- Dale, I., Clark, C., Stern, R., Leidi, S., & Stern, D. (2010). E-learning of statistics in Africa. *ICOTS 8* . Ljubljana.
- De Giusti, G. (2017). Using Digital Tools to Engage Kenyan Development Students with Data=. *Satellite Proceedings. Rabat: IASE*.
- Fundi, M., Clements, L., Stern, D., Stern, R., Renaud, F., & Sananka, A. (2017). Describing Data Well in R-Instat. *IASE Satellite 2017*.
- Fundi, M., Parson, D., Clements, L., Ndung'u, S., Sananka, A., Stern, D., et al. (2017). Key Features and Educational Uses of the Describe Menu in R-Instat. *IASE Satellite 2017*.
- Lunalo, J., Fundi, M., Kogo, S., Parsons, D., Clements, L., Ndung'u, S., et al. (2017). Key Features and Educational Uses of the Prepare Menu in R-Instat. *IASE Satellite 2017*.
- Manyalla, B., Mbasu, Z., Stern, D., & Stern, R. (2014). Measuring the Effectiveness of Using Computer Assisted Statistics Textbooks in Kenya. *ICOTS9*. Arizona.
- Musyoka, J., Lunalo, J., Garlick, C., Ndung'u, S., Stern, D., Parsons, D., et al. (2017). Embedding Data Manipulation in Statistics Education. *IASE Satellite 2017*.
- Musyoka, J., Stern, D., & Stern, R. (2014). Building Strength from Compromise: A Case Study of Five Year Collaboration Between SSC, University of Reading UK and Maseno University, Kenya. *ICOTS9*. Arizona.
- Odhiambo, J. W. (2002). Teaching of Statistics in Kenya. *ICOTS 6*. Cape Town.
- SSC. (2015). Statistical Services Centre. Retrieved from stats4sd.org/resources/e-learning
- Stern, D. (2014). Reflections on using technology to teach statistics in Kenya. *ICOTS 9*. Flagstaff.
- Stern, D. (2017). Seeding the African Initiative. *IASE 2017*. Rabat.
- Stern, D., Ongati, N., Agure, J., & Ogange, B. (2010). Incremental Modernization of Statistics Teaching and Curriculum At Maseno University, Kenya. *ICOTS8*. Ljubljana.
- Stirling, D. (2010). Improving Lectures using CAST Applets. *ICOTS8*. Ljubljana.
- UN. (2014). Data Revolution. Retrieved from UN Data Revolution: <http://www.undatarevolution.org/>
- Wolfram, C. (2010). Teaching Kids Real Math with Computers. Retrieved from https://www.ted.com/talks/conrad_wolfram_teaching_kids_real_math_with_computers