Teaching Experiences with a Course on "Web–Based Statistics"

Jürgen Symanzik
Utah State University
Dept. of Mathematics & Statistics
Logan, UT 84322–3900
e–mail: symanzik@sunfs.math.usu.edu

Natascha Vukasinovic
Utah State University
Dept. of Mathematics & Statistics
Logan, UT 84322–3900
e–mail: vukasino@sunfs.math.usu.edu

Abstract

Many Statistics courses have been taught that make use of Web–based statistical tools such as teachware tools, electronic textbooks, and statistical software on the Web. However, to our best knowledge, there has been no course before where statistical issues and the Web have been discussed systematically. In this paper, we provide an overview on our "Web–Based Statistics" course aimed at advanced undergraduate and beginning graduate students, including detailed discussions of lecture topics, homework assignments, and student projects. We discuss references (papers and URLs) useful for such a course and summarize students' feedback. We finish this paper with recommendations for future similar courses.

Keywords

Electronic Textbooks, Internet, Introductory Statistics, Teaching, Teachware Tools, WWW.

1 Introduction

Statistics and the Web become more and more interwoven these days. Examples are data sources on the Web, data collection and data distribution through the Web, on–line journals, teachware tools, and electronic textbooks on the Web. Also, many statistical software packages are freely available and immediately usable from within a Web browser. Incorporating software, in particular teaching software (or teachware) into introductory statistics courses has almost become a standard over the past decade. Using the Web in teaching statistics varies from occasional interactive exercises and in–class software demonstrations to completely Web–based courses without any use of "traditional" teaching tools or even pen and paper, such as the courses offered by Kent L. Norman at the University of Maryland (`http://cognitron.umd.edu/cognitron.html`).

However, individuals interested in statistics–related information available on the Web as well as those wishing to use available electronic teaching tools in their own introductory statistics class might find it difficult to grasp the amount and variety of the statistics–related material on the Web, even though papers such as Murdoch (2000), Symanzik (1998), Taerum & Nelson (1997), and Würländer (1997) try to provide a general overview on such material. It might be even more difficult to obtain information on where to find statistics–related sites and how to use the available material. Currently, students are occasionally exposed to a small portion of the existing Web–based statistical material, e.g., in an introductory statistics course that makes use of teachware on the Web or that uses a Web–based statistical software package. But prior to this course, there existed no course at Utah State University that systematically provides a general overview on Statistics and the Web, that introduces and compares several Web–based statistical software packages, and that makes todays students (and tomorrows teachers) aware of a multitude of teachware tools and electronic textbooks on the Web. The aim of our course on Web–based Statistics was to provide a remedy of this situation.

This paper gives an overview of our "Web–Based Statistics" course and our experiences gained through teaching this course. In depth details such as a complete list of all papers handed out in class or a breakdown of in–class activities and homework assignments can be found in Symanzik & Vukasinovic (2001) or directly at the course Web page at

`http://www.math.usu.edu/~symanzik/teaching/2000_stat5810/stat5810.html`.

In Section 2 of this paper, we discuss the motivation for the "Web–Based Statistics" course and its goals and provide an overview of the topics, in–class activities, homework assignments, group projects, and paper and Web–based references discussed in class. In Section 3 we summarize students' feedback related to the content, quality, and usefulness of this course. Recommendations for future "Web–Based Statistics" courses are given in Section 4.

2 The Course

2.1 Motivation and Goals

In May 2000 we taught an experimental "Web–Based Statistics" course for the first time at Utah State University's Department of Mathematics and Statistics. The course was aimed at advanced undergraduate students and beginning graduate students in Statistics and related fields such as Mathematics, Education, Computer Science, Administrative Sciences, and Natural Resources, and everyone interested in Statistics and the Web. Except basic statistical knowledge and computer, especially Internet literacy, no particular prerequisites were required.

The main goal of the course was to provide a broad overview of a large number of topics and motivate students to continue with further reading and to explore Web sites on their own. Therefore, the course was not intended to provide an in–depth discussion of a single topic such as detailed discussions on complex Web–based statistical software packages such as Rweb (Banfield 1999) and XploRe (Klinke & Kötter 1996, Schmelzer, Kötter, Klinke & Härdle 1996, Kötter 1996) or on electronic textbooks such as Hyperstat (Lane 1999) and the UCLA Statistics Electronic Textbook (de Leeuw 1997). Nevertheless, these topics still have been a major component of this course.

2.2 Course Format

The course was held at Utah State University's Department of Mathematics and Statistics in May 2000 during the first four–week Summer session immediately after the Spring semester. The course consisted of 15 lectures at 110 minutes each. Seven students registered for this course and two staff/faculty members audited.

The course took place in an electronic classroom where each student had access to a PC. In addition, the instructor's PC was connected to a projection device and, thus, the screen content of the instructor's PC was made available to the entire class.

There were three main formats for the lectures:

Joint Discussion: The topic was introduced at the beginning of a lecture and then students and instructors jointly discussed the issue. Such topics were the advantages and disadvantages of electronic textbooks or a discussion on which information a commercial software Web site should provide to its customers.

In–Class Activities: Students, working individually or in pairs, were given about 15 to 30 minutes (depending on the task) to explore a particular Web site. At the end of the given time period, students had to provide a short overview (3 to 10 minutes) of the visited Web site. However, due to time constraints, it was not always possible to provide all overviews during the same lecture, such that some presentations took place at the beginning of the following lecture. The main idea behind these activities was to effectively explore a Web site in a fixed time period rather than to get lost deep within a Web site. The five activities that were part of the course focused on statistical summary pages, data sources on the Web, teaching tools on the Web, electronic textbooks, and electronic journals on the Web.

Student Lectures: Students (individually or in groups of two) were given a few days to prepare a particular topic and then spend 20 to 45 minutes on teaching this topic to the entire class. Such lectures included "Introductory Statistics"–like segments that make use of a Web–based teaching tool or the presentation of a Web–based statistical software package.

In addition, a course Web page containing relevant information and all course–related material was created and regularly updated. This course Web page served several purposes: general announcements were made and homework assignments were posted on this Web page. No printouts of homework assignments were provided in class. Students had to submit the completed assignments in electronic form, typically by e–mail, and the answers have been added to the course Web page. This allowed the students who did not visit a particular Web site (or read a particular paper) to get a quick idea what this Web site (or paper) is about. The course Web page was also used to list an overview of discussed URLs and papers. It can still be accessed at

`http://www.math.usu.edu/~symanzik/teaching/2000_stat5810/stat5810.html`.

Grades were based on the quality of the in–class activities, the student lectures, the homework assignments, and the group project. No written exams were given.

2.3 Topics by Lecture

The major components of the course were related to Web–based teaching (Lecture 3 through 7) and statistical software on the Web (Lecture 8 through 12). The list of URLs discussed during the lectures is accessible through the course Web page. These URLs were not intended to provide a complete overview of all URLs related to a topic. Instead, they were selected as specific examples to compare and contrast different Web sites. The following topics have been discussed during the 15 lectures:

Lecture 1: "Introduction"

The introduction included an overview of the topics discussed in the course, an explanation of the course syllabus, an initial student survey, and an introduction to search engines on the Web.

Lecture 2: "Statistical Summary Pages & Data Sources on the Web"

Summary pages within major statistical organizations and universities, e.g., the American Statistical Association (`http://www.amstat.org`) and StatLib (`http://www.stat.cmu.edu`), and sources of statistical data e.g., FedStats (`http://www.fedstats.gov/`), have been presented and discussed.

Lecture 3: "Web–Based Teaching Tools"

Collections of applets related to teaching, e.g., The GASP Initiative (http://www.stat.sc.edu/rsrch/gasp/), have been presented in this lecture.

Lecture 4 & 5: "Electronic Textbooks"

These two lectures focused on a variety of Web–based electronic textbooks that are freely accessible. Examples are Hyperstat (http://www.davidmlane.com/hyperstat) and the UCLA Statistics Electronic Textbook (http://www.stat.ucla.edu/textbook/). Related printed references are Lane (1999) and de Leeuw (1997).

Lecture 6: "Web–Enhanced Teaching Statistics"

This lecture was focused on course Web pages (Leon & Parr 2000), their advantages and disadvantages, examples of course Web pages, and a survey on what undergraduate students thought of a particular course Web page.

Lecture 7: "Electronic Lectures"

This lecture consisted of student presentations of segments of an imaginary "Introductory Statistics" course that were enhanced by Web–based teaching tools (Lecture 3) or electronic textbooks (Lecture 4 & 5). Students could initially choose among seven different topics, prepare their lectures, and spend about 18 to 20 minutes on "teaching" the material. Topics chosen for the "lectures" were: Basic Graphics, Summary Statistics, Confidence Intervals and Hypothesis Testing, Probability, and Linear Regression.

Lecture 8: "Statistical Software on the Web"

An evaluation of (general) statistical software Web sites such as the SAS (http://www.sas.com) and S–Plus (http://www.insightful.com) sites has been conducted.

Lecture 9: "Commercial Software on the Web"

This lecture focused on possible directions for commercial software on the Web and a discussion on the advantages and disadvantages of accessing commercial software directly through the Web. A midterm survey was also conducted in this lecture.

Lecture 10, 11 & 12: "Software on the Web"

Three lectures were dedicated to statistical software on the Web. In Lecture 10, an introduction to Rweb (`http://www.math.montana.edu/Rweb`) and an overview of data analysis tools from the Rice Virtual Lab in Statistics (`http://www.ruf.rice.edu/~lane/rvls.html`) have been given by the two instructors. The following two lectures were student presentations. Lecture 11 consisted of an introduction to WebStat (`http://www.stat.sc.edu/webstat/`) and an introduction to Statlets (`http://www.statlets.com/statletsindex.htm`). Lecture 12 gave an introduction to XploRe (`http://www.xplore-stat.de`). Printed references related to these software packages were Banfield (1999) for Rweb, Lane (1999) for the Rice Virtual Lab, West & Ogden (1998) and West, Ogden & Rossini (1998) for WebStat, and Klinke & Kötter (1996), Schmelzer et al. (1996), and Kötter (1996) for XploRe.

Lecture 13: "Electronic Publishing"

The advantages and disadvantages of electronic publishing have been discussed, motivated by Buja, Heinke, MacEachren, Rossini & Cook (1997). Web sites of statistical journals that ranged from providing a short overview of the journal, e.g., The American Statistician (`http://www.amstat.org/publications/tas/`) and Computational Statistics (`http://comst.wiwi.hu-berlin.de/`), to full electronic journals, e.g., the Journal of Statistics Education – JSE (`http://www.amstat.org/publications/jse`) and the Journal of Statistical

Software – JSS (http://www.jstatsoft.org/), have been visited.

Lecture 14: "Electronic / Web–Based Data Collection"

Questions in this lecture were who might be interested in Web–based data collection and why. Answers were given in articles such as Lee, Shing & Chu (1996), Sear (1997), and Keller & Bethlehem (2000). Advantages and disadvantages of Web–based data collection were also discussed.

Lecture 15: "Web–Based Data Distribution"

The final lecture dealt with the question who might be interested in Web–based data distribution (e.g., the Federal Government) and existing Web–based applications to allow an interactive access to the data (rather than to download the data in a simple spreadsheet format), e.g., the Graphics Production Library – GPL, accessible at http://www.monumental.com/dan_rope/gpl/ and further described in Carr, Valliant & Rope (1996).

2.4 Homework Assignments

Homework assignments were given approximately three times a week. In total, there were 12 homework assignments. Homework assignments comprised three main types of assignments. The first type of homework assignments was related to the in–class activities. For each of the five activities described in Section 2.2 and for some of the visited Web sites, the students had to turn in a brief written summary as a homework assignment. The second type of homework assignments aimed at preparing the students for the in–class presentations. The students were also asked to evaluate the other students' presentations. The results were anonymously summarized and provided to the presenters. In the third type of homework assignments, the students were asked to summarize one paper of their choice on a particular topic. Typically, three or four papers were available for each of the topics.

The students had to submit answers to the homework assignments electronically and these asnwers were posted on the course Web page. This allowed every participant in the class to see what other participants thought of a Web site or a paper they had not visited or read themselves. Evaluations of student presentations were not posted on the course Web page and were only made available to the student(s) who presented a particular short lecture or a software tool.

A detailed overview of all 12 homework assignments can be found in Symanzik & Vukasinovic (2001). Also, the complete set of assignments turned in by the participating students is still available at the course Web page at

`http://www.math.usu.edu/~symanzik/teaching/2000_stat5810/homeworks/homeworks.html`.

2.5 Group Projects

In addition to the regular course work, the students had to work on a major group project dealing with a topic related to the content of the course. The group projects were included as a part of the course to compensate for the smaller number of lectures and the reduced overall course time in the summer session. The students had to provide an in–class presentation of their work and a written report.

One group of two students worked on the numerical performance of Web–based statistical software packages (i.e., WebStat and Statlets) using the Statistical Reference Data Sets (StRD) from the National Institute of Standards (NIS), closely following McCullough (1998), McCullough (1999*a*), and McCullough (1999*b*). The related project report has been enhanced to a journal paper (Kitchen, Drachenberg & Symanzik 2002) over the following months and has been submitted for publication in Computational Statistics.

The other group of four students worked on the comparison of electronic textbooks for possible use in undergraduate statistics courses. We are still working on the enhancement of the project report into a publishable journal paper.

2.6 Course Material

A large number of papers has been distributed to the students during the course and many Web sites related to Statistics and the Web have been visited. It was not expected that students read all of the papers. In particular, for some of the homework assignments students had to choose one paper that looked most interesting to them and summarize that paper.

Main sources of papers were the "Computing Science and Statistics" proceedings from the last few years. Additional material has been found in major conference proceedings such as the "CompStat" proceedings and the "Bulletin of the International Statistical Institute". Very little material has been found in statistical journals. Due to this fact, it is highly recommended to plan well ahead to obtain the desired reading material because many libraries only hold some, but not all of the conference proceedings. Obtaining a copy of a desired conference paper through interlibrary loan often takes a while.

2.6.1 Papers

Paper references consisted of papers published in conference proceedings or journals, abstracts of upcoming talks, technical reports, preprints, or draft versions of papers. We also speak of a "paper" when a single document, e.g., an HTML, a PostScript or a PDF file, has been obtained from a Web site. Over time, it should be possible to get the final version of those references that were not yet a published paper at the time this course was held. Due to the large number of publications in conference proceedings on Statistics and the Web, it should be possible to easily replace some of the older references by current material in a future offering of this course.

A complete list of paper references used in the course can be found in Symanzik & Vukasinovic (2001) and is available on the course Web page at

`http://www.math.usu.edu/~symanzik/teaching/2000_stat5810/web_refs_topic.html`.

2.6.2 URLs

All URLs that have been visited during this course are listed on the course Web page at

`http://www.math.usu.edu/~symanzik/teaching/2000_stat5810/web_urls_may2002.html`.

The URLs have been ordered by topic and have been updated in May 2002. If a URL no longer exists, it should be possible to locate a similar URL for a future "Web–Based Statistics" course.

3 Summary of Student Feedback

During the course, we carried out two informal surveys to assess students' expectations as well as their opinion about the ongoing course. At the end of the course, the official University Teacher/Course Evaluation was conducted. About 15 months after the course ended, a post–course survey has been conducted to assess the long–term impact the course had on the students and their attitudes toward the use of Web–based statistics–related material.

In summary, most of the students started off with just basic experience in Web browsing and did not really know what to expect of a "Web–Based Statistics" course. However, as the course advanced, they were surprised to find out how much statistics–related material was available on the Web. Interaction, in–class presentations, hands–on experience on exploring the Web and, in particular, group projects, were the most appreciated aspects of the course. Generally, the students were satisfied with the content of the course and the topics that were discussed. Some students complained about heavy workload and the dense course schedule in the 4–week summer session. Instead, they would prefer taking this course in a regular semester. Finally, it turned out that the most useful Web sites discussed in the course were those related to statistical software on the Web, followed by data sources, statistical summaries, and statistical journals on the Web.

4 Conclusions

Teaching the "Web–Based Statistics" course provided valuable experiences for both the instructors and the students. In general, it can be concluded that 15 lectures at 110 minutes each is too little time to allow for an in–depth discussion of all aspects of Statistics and the Web. However, this is enough time to give a general overview of the material. If more time was available, e.g., during a regular semester, topics such as Web–based software and Web–based teaching tools could be easily extended and discussed in more detail.

Depending on the students' backgrounds and interests, additional topics could be added to the course. Examples of possible topics are the basic coverage of creating Web pages, in particular HTML, advanced techniques of Web design such as Javascript, Java, xml, and Flash, or analysis and visualization of Web traffic or electronic commerce. Based on the backgrounds and interests of students attending future "Web–Based Statistics" courses, it could easily be possible to adjust the topics to make the material presented and discussed in the course interesting for a variety of majors.

Based on the results of the three informal surveys and the official course evaluation, it can be concluded that the students liked both the content and the format of the "Web–Based Statistics" course and that they were surprised to see how much material related to Statistics can be found on the Web.

One of the main recommendations for future "Web–Based Statistics" courses would be to run the course during the regular semester. This would certainly allow for more time to work on homework assignments and projects. In a full semester it would also be possible to discuss some of the topics in more depth or add additional topics as outlined above. Overall, students liked the course format with the in–class activities and presentations, so that the format of the course should be maintained in the future.

Acknowledgments

We would like to thank Bruce Bayley, Weiping Deng, Roxanne Drachenberg, Glen Erickson, Ann Kitchen, Nancy Roberts, Gary Straquadine, Jingjing Wang, and Qian Zhao, i.e., the students and staff/faculty members from Utah State University that participated in this course, for their helpful comments and suggestions.

References

Banfield, J. (1999), 'Rweb: Web–based Statistical Analysis', *Journal of Statistical Software*. http://www.jstatsoft.org/v04/i01/.

Buja, A., Heinke, R. S., MacEachren, A., Rossini, A. J. & Cook, D. (1997), Publishing in the Electronic Age, *in* '1997 Proceedings of the Section on Statistical Graphics', American Statistical Association, Alexandria, VA, pp. 1–5.

Carr, D. B., Valliant, R. & Rope, D. (1996), 'Plot Interpretation and Information Webs: A Time–Series Example from the Bureau of Labor Statistics', *Statistical Computing and Statistical Graphics Newsletter* **7**(2), 19–26.

de Leeuw, J. (1997), The UCLA Statistics Textbook and Modules, *in* 'Bulletin of the International Statistical Institute, 51st Session Istanbul 1997, Proceedings Book 2', pp. 55–58.

Keller, W. J. & Bethlehem, J. (2000), 'The Impact of EDI on Statistical Data Processing', *Computational Statistics* **15**(1), 53–63.

Kitchen, A. M., Drachenberg, R. & Symanzik, J. (2002), 'Assessing the Reliability of Web–Based Statistical Software', *Computational Statistics*. Submitted.

Klinke, S. & Kötter, T. (1996), XploRe 4 — A Statistical Computing Environment, *in* F. Faulbaum & W. Bandilla, eds, 'SoftStat '95 Advances in Statistical Software 5', Lucius & Lucius, Stuttgart, pp. 113–122.

Kötter, T. (1996), Development of XploRe 4: The Programming Language, *in* A. Prat & E. Ripoll, eds, 'Compstat – Proceedings in Computational Statistics, Short Communications and Posters', pp. 201–202.

Lane, D. M. (1999), 'The Rice Virtual Lab in Statistics', *Behavior Research Methods, Instruments & Computers* **31**(1), 24–33.

Lee, T.-W., Shing, C.-C. & Chu, S.-C. (1996), 'Automating Statistics in WWW', *Computing Science and Statistics* **27**, 485–489.

Leon, R. V. & Parr, W. C. (2000), 'Use of Course Home Pages in Teaching Statistics', *The American Statistician* **54**(1), 44–48.

McCullough, B. D. (1998), 'Assessing the Reliability of Statistical Software: Part I', *The American Statistician* **52**(4), 358–366.

McCullough, B. D. (1999*a*), 'Assessing the Reliability of Statistical Software: Part II', *The American Statistician* **53**(2), 149–159.

McCullough, B. D. (1999*b*), 'Experience with the StRD: Application and Interpretation', *Computing Science and Statistics* **31**, 16–21.

Murdoch, D. J. (2000), 'Statistics on the Web', *Chance* **13**(1), 49–51.

Schmelzer, S., Kötter, T., Klinke, S. & Härdle, W. (1996), A New Generation of a Statistical Computing Environment on the Net, *in* A. Prat, ed., 'Compstat – Proceedings in Computational Statistics', Physica–Verlag, Heidelberg, pp. 135–148.

Sear, J. (1997), Electronic Reporting and Its Use for Business Surveys, *in* 'Bulletin of the International Statistical Institute, 51st Session Istanbul 1997, Proceedings Book 2', pp. 449–452.

Symanzik, J. (1998), 'Current Internet Technology and Statistics — Blessing or Curse?', *Computing Science and Statistics* **30**, 500–509.

Symanzik, J. & Vukasinovic, N. (2001), 'Experiences with a Course on "Web–Based Statistics"', *Computing Science and Statistics* **33**, CD, Forthcoming.

Taerum, T. & Nelson, T. (1997), 'Internationalization of Statistics: The Internet', *Computing Science and Statistics* **29**(2), 354–359.

West, R. W. & Ogden, R. T. (1998), 'WebStat: An Environment for Statistical Analysis on the the World Wide Web', *Computing Science and Statistics* **29**(1), 307–310.

West, R. W., Ogden, R. T. & Rossini, A. J. (1998), 'Statistical Tools on the World Wide Web', *The American Statistician* **52**(3), 257–262.

Würländer, R. (1997), The Internet Link: Resources and Access to Research, *in* 'Bulletin of the International Statistical Institute, 51st Session Istanbul 1997, Proceedings Book 2', pp. 51–54.