Abstract

The Internet and the WWW have brought a new dimension to statistical education, given the availability of a variety of resources. These may be used as primary or supplementary material and range from comprehensive electronic statistical and methodological textbooks to applets that demonstrate a single statistical idea. They include specialized guides, demonstrations, and data sets. Online materials may make use of hypertext, graphics images, sound files, or text. Many instructors place their course syllabus, assignments, PowerPoint presentations, and Web links online for students. Instructors can ask students to search the archives in e-journals or to interact with each other via online bulletin boards or computer conferencing. An integral part of distance education, such tools can provide tremendous enhancement to the traditional classroom course. Given the range of options, how easy is it to find the proper resources for inclusion in a statistics course? How much material is available? How do we evaluate it? What resources exist to help faculty use the Internet resources well? These questions are addressed and, while some useful sites are provided, the main goal is to provide a discussion about the issues of finding, evaluating, and organizing Internet resources.

Keywords: Statistics teaching; Finding and organizing Web resources; Evaluating online resources

1. Introduction

The introduction of the Internet and the World Wide Web (WWW) has opened new possibilities and directions for teaching in general and for the teaching of statistics in particular. The availability of such tools has helped transform the way statistics is taught. If not an actual reality, this is certainly a belief shared by many statisticians and methods instructors. The use of some form of interactivity via the WWW is encouraged for statistical instruction from university down to school level. The focus on adding at least some online components to courses is based on the perceived utility to students of such additions. Of course, this is true across disciplines, and not only for statistics curricula. At my own university, an on-going program is in place that provides technology training for faculty with the goal being to motivate them to learn about and use alternative instructional strategies while providing access to state-of-the-art instructional technology. This program supports and encourages faculty to develop network accessible course materials and electronic libraries supporting select courses. While not mandated, such use is very strongly encouraged across disciplines.

Whether used as part of classroom lectures and demonstrations, as supplementary materials for homework, or as part of distance learning courses, we have accepted the advantages of using this technology for delivery of course material over more traditional printed formats. For example, Schwarz and Sutherland (1997) provided the following list of advantages to delivering course material via the WWW:
• “It can deliver a mixture of media (text, graphics, programs).
• Instructors can customize material easily.
• Workshops can be updated quickly and easily, and updates are immediately available.
• It is suitable for distance education courses.
• It has a “just-in-time” delivery that avoids problems of ordering and storing printed materials.
• It can be enriched with hyperlinks to supplemental material.”

It is difficult to argue with these advantages. But, how easy is it to locate the desired information and how much material is available for use in statistical instruction? Once found, how do we evaluate it? What resources are available to help faculty customizing, enriching, or updating their courses with electronic resources? These questions will be addressed and used as places to start a discussion about issues surrounding statistical education on the Internet.

2. Locating Information

In their report on online teaching, Lee, Armitage, Groves, & Stephens (1999) stated: “The World-Wide Web (or the Web) is one of the most accessible tools available for academics to use. It allows an easy means of publishing material, it has a low learning-curve, the majority of its browsers are graphical and user-friendly, and above all it is free to most people in Higher Education” (Chapter 4). This tool fits well with the direction of student-centered resource-based learning. This ever-expanding resource, however, presents new problems of locating the appropriate information. The report cites two main approaches for solving this problem: (1) Internet gateways, which are lists of links organized by subject and sometimes annotated, and (2) Internet search engines, which use computer query languages to search out and index web pages based on key words the user provided. The authors note that gateways are a useful starting point because, even though they are not and cannot be exhaustive in their coverage, they are typically run by “an enthusiast, a subject expert, or, increasingly, a librarian” who then “bring their judgment to bear on the sites they link to” (Lee, Armitage, Groves, & Stephens, 1999). For certain specific topics or interdisciplinary searches, however, the selectivity of gateways may not be productive and the greater reach of search engines is needed.

All six gateways that are listed in that document are hosted by the United Kingdom. Of those, perhaps the most useful for statistical teaching information is the Social Science Information Gateway (URL: http://sosig.ac.uk/). The statistics link goes to demographic and official statistics lists and sites. From general social science, one can go to social science methodology and then to psychometrics and statistics and methodology, with links to individual as well as collections of articles, papers, and reports, bibliographic databases, books or book equivalents, educational materials (including one about informed consent for research), two lists of journals (one with contents and abstracts only and one with full text), mailing lists and discussion groups, organizations and societies.

The five search engines listed (AltaVista, HotBot, InfoSeek, Excite, and Lycos) do not include Google and AlltheWeb, two recently top-rated search engines (Notess, 2003). These ratings were found in a link from a site called “Search Engine Showdown: The User’s Guide
to Web Searching” (Notess, 1999/2003) that provides information such as reviews and statistics, a search engine features chart, and points to sources on how to learn about searching and search engines and how to use them well. As will be demonstrated later in the paper, this ability is crucial for conducting profitable searches.

3. Breadth of Statistical Education Information Online & Search Engine Differences

One can quickly get a sense of the magnitude of information on the Web. Looking first to issues about teaching statistics, I ran searches on AltaVista, Google, and Yahoo using a few general terms. Across these three search engines, the term statistics education brought up between 2.4 and 2.9 million hits, while statistics instruction produced from 444 thousand to more than one million in the different searches. The resulting hits dropped dramatically when I put quotation marks around the terms, bringing the numbers down to between 1,000 and 18,000. The phrase “statistics teaching” averaged 2,000 hits across the three sites, while its inverse “teaching statistics” averaged 11,000. AltaVista provided some suggestions for refining the searches, such as: introductory statistics, mathematical statistics, teaching resources, teaching statistics, Java applets, statistical analysis, datasets, data analysis.

In searches for “statistics instruction” on five search engines (AltaVista, Google, Yahoo, Lycos, and Hotbot) the first link was to “MAA99 – Using Web Applets to Assist Statistics Instruction,” a page by Robin Lock from a 1999 presentation at the Joint Mathematics Meetings, which provides links to student activities, other websites with links to applets, and online statistical packages, as well as a sample course page. Although there were several other common links, each search engine also brought up different sites on its first screen.

The term “statistics teaching” produced only a few of the same links as did “statistics instruction.” A new listing was the “Statistics Teaching Resources and Related Links” page (http://www.niss.org/copss/teaching.htm), a gateway that provides links under three topics: general resource links, statistics courses on the web, and statistics education. This was contained under the Committee of Presidents of Statistical Societies page (COPSS), which includes the American Statistical Association, the International Biometric Society, and the Statistical Society of Canada and is maintained by the National Institute of Statistical Sciences.

What is striking is the serendipity of the first screen results. While numerous relevant sites were found, Yahoo brought up an empirical article on “Cooperative learning and statistics education” from the 1997 Journal of Statistics Education, but failed to bring up any other comparable or more recent articles. Two links from a Lycos search both related to an article in The American Statistician by Lovett and Greenhouse (2000) entitled “Applying cognitive theory to statistics instruction.” One link was to a Scientific Literature Digital Library (CiteSeer), which is a research index page (http://citeseer.nj.nec.com/239583.html) and the other was to an American Statistical Association page (http://www.amstat.org/publications/tas/2000/Lovett.htm). In both cases, the links contained an abstract of the article. What is interesting is that a full text version of the article is available online (http://www.stat.cmu.edu/www/cmu-stats/tr/tr689/tr689.html), but this search did not bring it up, at least not in the first screen of results.
One resource site produced in the first 10 hits by a search engine ceased to exist in 1999, but the link forwards the user to another resource: “Maths, Stats, and OR Network,” which is hosted by the University of Birmingham in partnership with Glasgow and Nottingham Trent Universities and the RSS Centre for Statistical Education (http://ltsn.mathstore.gla.ac.uk/index.asp?view=sor). One search engine’s top 10 results was a link related to a 1992 faculty project for an introductory course for health, psychology, and education students at Pennsylvania State University and another simply contained images like dice and cards to be used in spicing up presentations on probability. Other hits in the first page of search results included a link to an online journal (Computational Statistics), several statistical reports from different Departments of Public Instruction in several states in the USA, and some individual web sites devoted to statistics.

Using more specific phrases in quotes resulted in relatively few hits. For example, searching with AltaVista, Google, and Yahoo with “teaching statistics online” produced 19, 33, and 35 sites, respectively, while “web resources for teaching statistics” produced only 8, 11, and 24, respectively. No sites were found using the altered phrase “statistics teaching online.” This points to the utility of both using different search engines and different keywords.

Turning to searches of specific statistical topics, a few examples show the general magnitude of available information from which to choose. For example, searches for “multiple regression” produced up to 198K links, while for “structural equation modeling” the number fell to 97K. Even a relatively simple term such as “Pearson correlation coefficient” produced up to 8K hits one month and over 9K two months later. Dropping the name Pearson increased the hits to over 70K.

Why do such seemingly dramatic differences exist in results across search engines and keywords used? One answer is that search engines follow computer database searching rules, most of which are based on Boolean logic, but they don’t do so in an identical manner. One way they differ is in the use of full Boolean searching, which requires the use of the Boolean logical operators (AND, OR, NOT), or implied Boolean logic with keyword searching. In the latter case, symbols replace Boolean logical operators, but the absence of a symbol also carries meaning. For example, the space between keywords may default to OR logic or AND logic, depending on the search engine. That means that searching using the keywords statistics education could result in a search for any occurrence of either term appearing in a document or for cases where both terms must occur. According to Cohen (2003) “Many well-known search engines traditionally defaulted to OR logic, but as a rule are moving away from the practice and defaulting to AND.” Not knowing which is the default could certainly cause problems. This is increasingly so because “Implied Boolean logic has become so common in Web searching that it may be considered a de facto standard” (Cohen, 2003). Putting quotes around a phrase implies that the exact phrase must appear. But, using “statistics education” would not find sites with “statistical education” or other variants on the term.

The University at Alabama Libraries offers a series of Internet tutorials (Cohen, 2003) that include selecting a tool for an Internet search, a primer in Boolean logic, selecting a search engine or directory, and recommended sites and search techniques (see reference list for URLs for main site as well as for each tutorial). Useful information about web searching
may also be found in Notess (1999/2003) and, for details on specialized terminology, see “Terms relating to database and Internet searching” (Hansen, 1998/2003).

4. Evaluating Web Sites

Given the vast, dynamic, and ever-expanding nature of the Internet and the presumed wide range in the quality of information available, it is essential that both teachers and students evaluate that information. A search of the Web with keywords such as “website evaluation” is one place to start. Some examples of sites devoted to information about how and why to evaluate web resources and web information are provided by way of illustration of available resources in this area. Although by no means exhaustive, they were chosen because they were often cited, they provide additional resources, and are relatively current.

The University Libraries at Virginia Tech maintain a web page with an extensive bibliography of “documents which address the problems and issues related to teaching and using critical thinking skills to evaluate Internet resources” (Auer, 1997). The site has a link to Tech’s “Interactive Module on Evaluating Internet Resources,” which is an interactive modification of Susan Beck’s evaluation criteria page. Beck (1997), the Head of the Reference and Research Service Department of the New Mexico State University Library listed five criteria for evaluating web sources: accuracy, authority, objectivity, currency, and coverage in her website. She provides both useful descriptions of what to look for and a rationale for each criterion. She also provides suggestions to instructors for successful internet assignments such as providing guidance, checking the site to make sure it is still functional, making sure students record time and date of access, and using established, trustworthy web guides and directories, among others. The Virginia Tech site includes links to web pages as examples of each criterion.

Schrock (1995) created a set of materials to help students “critically evaluate a Web page for authenticity, applicability, authorship, bias, and usability.” She designed a series of checklists for use at the elementary, middle, and secondary school levels in both English and Spanish, which are available in both html and pdf formats on her website. In one revised article, she provided 26 criteria for evaluating Web pages. Although both general and commonsensical in nature, her suggestions are a practical guide that students at all levels could find useful as they navigate the web. Among other things, she suggested evaluating the author’s credibility, looking for a source bibliography, and verifying the electronic information via reputable print sources. She also suggested evaluating a site’s scholarliness by reviewing what sites have linked to it, saying, “This type of Internet search can be conducted using HotBot and AltaVista. [http://www.hotbot.com/ and http://www.altavista.com/]. HotBot employs a drop-down menu to search for links to individual sites, while AltaVista allows users to type link:<URL> in its search box” (Schrock, 2002). For teachers, she stresses the importance of testing a site for download time at various times of day or asking for a site author’s permission to retrieve the site and make it available on your own computer with WebWhacker (http://www.bluesquirrel.com/whacher/) or WebBuddy (http://www.dataviz.com/products/webbuddy/).

In their web evaluation site, Alexander and Tate (1996) provide an instructional model with two goals: (1) “Provide materials to assist in teaching how to evaluate the informational content of Web resources”, and (2) “Provide a bibliography of materials on applying critical
thinking techniques to Web resources.” Their bibliography contains web evaluation articles and books as well as links to additional web evaluation sites, several of which deal with more scholarly issues such as the critical analysis of information sources, including journals. They also make available several evaluation instruments that they designed for different types of Web pages because they contend that different criteria are needed for different types of pages.

A page maintained by the WWW Virtual Library is aimed at librarians and others “who are selecting sites to include in an information resource guide, or informing users as to the qualities they should use in evaluating Internet information” (Smith, 1997). The resources are organized into three categories: (1) general selection criteria, (2) selection criteria used for specific sources, and (3) commentary.

5. Site Identification Issues & Other Search Related Problems

Although Web pages and search engines can be very useful tools, it is not only the massive amounts of information that can be daunting. For example, a top ten find in one search pointed to Outline: WWW Resources for Teaching Statistics (http://it.stlawu.edu/~rlock/tise98/outline.html), which provides organized sets of links for online course materials and texts, JAVA demonstrations, electronic journals and discussion lists, data, and miscellaneous links. However, it is not obvious from this page who created it or when. It is only after Section 9, the conclusion, that there is a link back to a page that identifies the author and his institution, as well as the fact that this was a paper presented in 1997 by Robin Lock and called Internet Resources for Teaching Statistics (http://it.stlawu.edu/~rlock/iase97/index.html). An inspection of the two URLs indicates that one is more recent than the other. In fact, the former is an updated and enhanced version of the latter, following exactly the same outline. In the 1998 version (Lock, 1998), the conclusion screen has a ’97 URL and the return link brings one back to the original 1997 “main outline” page with the author’s name and pertinent information about the paper. But now, if one opens a link from the main outline, one is in the older version of the webpage, with fewer listings. It is only a careful scrutiny of the URLs that makes the difference obvious. This could cause problems for students who were given the former site to use, arrived at the conclusion, then went back to the main menu to return to an earlier source list only to find a different set of links than the instructor expected them to see.

Another example shows how potentially useful statistical pages would inadvertently fall short based on any set of website evaluation criteria. Consider the following URL, which resulted from a search on the term statistics education: http://www.helsinki.fi/~jpuranen/links.html. The only information given in this page is “This is my list of some useful links related to Statistics Education – Online Statistical teaching material, courses, handouts, exercises, articles, datasets, Useful lists of links, etc.” It does provide a link to contact the author, but gives no information about the author or his/her affiliation. Although the list is quite extensive, it was last updated 2 December 2002. It does, however, provide a link to “Noppa5-links” (http://noppa5.pc.helsinki.fi/links.html), which is said to be a “usually more up to date list of links.” This is, in fact, a more recently updated (26 March 2003) list, but again provides no information about the author or with which to evaluate the page. By going to the first part of the URL, we find the homepage of Juha Puranen, Dept of Statistics at University of Helsinki. So, the information is findable, but one needs to deconstruct the URL in order to obtain it. The statistics education page is
certainly not without merit. It points to links that are organized according to the following topics: online statistical material, statistical texts, graphics, courses (including handouts, exercises, quizzes), statistical journals and articles, teaching materials, data sets, lists of other links, statistical demos, statistical software, organizations and lists, history, and teaching and teaching tips. Additionally, in a *Bulletin of the International Statistical Institute*, Saporta (1999) says about this site “J. Puranen has one of the most complete list of links for teaching material.” Yet, someone using a Web site evaluation checklist might not have rated the original page well.

Both of these examples speak more to the organization of websites than to their content. Part of the problem is the fact that search engines may produce pages that are secondary links, without identification, even when primary pages that have complete ID information are not brought up. We need to be aware of this when we create Web pages and learn how to better document each sub-page or, at least, include a link to a main identifying page on every portion of a web site. Additionally, we can instruct our students about deconstructing URLs to search out authors or sponsoring organizations for pages of interest. Another problem is likely due to the fact that older versions of pages remain on the Web when newer, updated versions are created. To Beck’s (1997) recommendation that students should record time and date of access, we should add that they should pay attention to the date of creation and/or modification for the pages they are using. Likewise, as instructors, we need to make sure of providing such dates for students to ensure that the material we think they are viewing is, in fact, what they are using.

Deconstructing URLs can be a useful strategy in other ways, as well. For example, there are numerous educational technology conferences with papers available online, which can provide useful current information. An online search produced links to over a dozen such sites, however, many of them had dates as old as 1999. In some cases, these links were to small one-shot conferences and, certainly, there were links to current websites for 2003 or even 2004 conferences. But, in other cases, the link was to a dated yearly conference, with no link to a more recent version. In such cases, one can sometimes change the year on the URL and get to a current listing for the conference or to find a paper from a different year’s electronic proceedings. In other cases, one can use the initial core of the URL to find a sponsoring agency or association page, and then link to conference and proceedings sites.

### 6. Faculty Support and Resources for Technology Tools

There is a world of possible options between an entirely online distance learning course and simply putting text-based course notes online for students to download. Finding existing materials that give guidance to what one could or should do as well as finding statistical sites to use in one’s courses is only part of the equation. What resources are available that provide a more formalized support system to learn how to enrich, update, or customize one’s courses and take advantage of the available electronic tools? I will give an example of what my own university is doing, and then provide examples of some general university resources.

The Instructional Development Initiative at Virginia Tech is an ongoing project that started with three pilot faculty workshops in 1993. To date, well over 200 faculty workshops have been conducted, numerous classrooms have been upgraded with computers and projection systems and courses restructured. This large-scale university-level initiative is
aimed at improving the teaching-learning process by providing faculty with “the opportunity to rethink their teaching and explore the potential of instructional technology” (Instructional Development Initiative, 2002). Assessment of the program, mostly via faculty and student surveys, has produced positive results. A 2002 report claims, “There is evidence that these efforts have had a positive impact on students’ understanding of and interest in the course material while promoting better class attendance. In addition, students believe they are being provided more opportunities to develop skills that transcend the subjects matter, including problem-solving and critical thinking” (IDI, 2002, Annual Report, pp.3-4).

Universities often maintain electronic sites aimed at facilitating instructional use of technology. Such sites can provide a wealth of resources for faculty. Three such university Web pages are:


- The School of Education at the University of Colorado at Denver has lists of resources, readings, and corollary sites for Teaching and Learning on the Internet ([http://carbon.cudenver.edu/~mryder/itc_data/net_teach.html](http://carbon.cudenver.edu/~mryder/itc_data/net_teach.html)).

- The University of Maryland University College’s Virtual Resource Site for Teaching with Technology, with modules for teaching/learning activities and online courses ([http://www.umuc.edu/distance/odell/vteach/module1/strategies.html](http://www.umuc.edu/distance/odell/vteach/module1/strategies.html) and [http://www.umuc.edu/distance/odell/vteach/module2/mod2.html](http://www.umuc.edu/distance/odell/vteach/module2/mod2.html)), is a joint venture of the Center for the Virtual University and the Center for Teaching and Learning.

Although our purpose is teaching statistics, understanding the media and how it may be used is an essential component to successful use and incorporation of technology into the classroom, which is why the sources like the above are both relevant and important. However, a few resources that can provide perspective on what is specifically available for statistics follow.

- One example of a gateway is the IASE’s “Resources for Teaching Statistics” page ([http://www.swin.edu.au/maths/iae/se/resources.html](http://www.swin.edu.au/maths/iae/se/resources.html)). While the page is not identified as an IASE page, it does provide a link to the IASE homepage as well as to Brian Phillips’ email. Links to additional statistical education sites can be found on the IASE homepage ([http://www.swin.edu.au/maths/iae/](http://www.swin.edu.au/maths/iae/)). These include statistical education discussion lists and research in statistical education, among others.

- Saporta (1999) provides a good survey of available resources for using the Internet in statistics teaching. He provides evaluative commentary along with the URLs for gateways (lists of links) maintained by both associations and individual researchers, for applets, free software, databases, electronic textbooks, distance teaching and the St@tNet project. (from [http://www.stat.fi/isi99/proceedings.html](http://www.stat.fi/isi99/proceedings.html); go to “invited authors” link and select Saporta’s article).

• Würländer, R (1997) The Internet link: Resources and access to research (http://ourlworld.compuserve.com/homepages/Rainer_Wuerlaender/).

Symanzik and Vukasinovic (2003) described a course at Utah State University that they believe is unique. It provides both a general overview of statistics and of use of the Web. The course was aimed at students in statistics and related fields, as well as others interested in statistics on the Web. Two main components of the course involved Web based teaching and statistical software on the Web. Such a course may be a useful way to acquaint future instructors to become adept at and make better use of the available technologies in their statistics classes. In their paper, they provided an overview of course content and a link to their course web page. (The URL for their course web page is http://www.math.usu.edu/~symanzik/teaching/2000_stat5810/stat5810.html). Their course was divided into 15 lecture, with two major components relating to Web-based teaching (5 lectures), statistical software on the Web (5 lectures). Other lectures dealt with search engines, summary pages and data sources, electronic publishing, electronic data collection, and Web-based data distribution.

7. Conclusion

Based on a review of the literature spanning 1983 to 2000, Mills (2002) concluded that there was very little empirical research to support the advocacy for computer simulation methods (CSM), even though the consensus was that they were effective. Only two of the 48 articles listed in her appendix were empirically based. The belief that such models will aid learning seems to be based on constructivism, a theory that holds that students learn by doing and that teachers can facilitate such learning with interaction and discussion. Although many of the articles she reviewed discussed “students involved in their own learning and construction of knowledge,” Mills (2002) found that only a few articles that “specifically mentioned or identified a general theory of learning.” The more recent of the two empirical studies she cited was by delMas, Garfield, & Chance (1999), an action research study with the goal of gaining insight into and improving their classroom instruction, initially in the area of statistical inference. They also concluded that “despite the accepted approach used to integrate simulation software into a statistics class, there is little published research describing and evaluating such an approach.” This is an often-heard complaint about various Web-based approaches for statistics instruction; that strategies or applications used are presented with little more than anecdotal evidence of merit.

In their exploratory investigation of three online delivery approaches, Edwards and Fritz (1997) took the approach that any question about effectiveness “more accurately refers to the effectiveness of ‘learning’ rather than ‘teaching’.” After summarizing a debate over what constitutes learning, they argued that student perceptions of their own learning are an important indicator, and this is what they assessed. Their results were generally positive, but the sample sizes were small and the comparisons confounded. Another argument often heard is that sound comparative educational studies are difficult to do. However, Schuyten, Dekeyser, and Goeminne (1999) did an experiment, with random assignment of about 500 students and appropriate controls, to examine the viability of an electronic independent learning environment. Although such studies are few and far between, they are always called for in the literature.
I am not arguing against empirical studies to demonstrate what is or is not effective with respect to statistical instruction with online resources. However, the reality is that the Internet and the Web do exist and the statistical community, university administrators, and students are demanding greater hands on experiences with statistics, which the Web can provide. So, without the requisite studies, how do we deal with the vast expanse of resources of potentially varying value? I would contend that we do so in the same way we deal with the relatively vast expanse of statistical text material such as textbooks and journal articles. We have editors and reviewers who help screen some material and we are selective in what we bring into the classroom. So too, we need to transfer the skills used to evaluate static resources and use them to make judicious choices in the interactive online world. An Assistant Dean of University Libraries for Reference and Information Services at the College of William and Mary in Virginia called the Internet the “proverbial firehose.” He cited numerous sources where the information is either scant or wrong, and argued that “more filtering of information by professional editors and publishers before it reaches the Internet will help put the squeeze on the firehose's flow, as will the guidance of reviews embodying criteria appropriate for this relatively new phenomenon” (Rettig, 1995).

There are and probably should be “many faces” of statistical education and many ways to effectively use Internet resources as an integral part of our courses. In this paper, I hoped to provide a few examples of different resources that can help us make better use of the statistical information that we find online. We know the content; we need to better understand the process of finding, evaluating, and organizing electronic statistical resources.

REFERENCES

(NOTE: All electronic references were retested 21 June 2003)


3. “How to choose a search engine or directory” (3 June 2003) http://library.albany.edu/internet/choose.html
4. “Searching the Internet: Recommended sites and search techniques (7 May 2003) http://library.albany.edu/internet/search.html


