Supplementary materials for this article are available online. Please click the TAS link at http://pubs.amstat.org.

Statistics in the News

Jean-François PLANTE and Nancy REID

We describe the structure, topics, and activities used in teaching a first-year seminar course in statistical literacy. This course is entirely structured around current affairs, although common themes emerge that reflect the core material for most quantitative literacy courses. Special emphasis is given to the computing and calculating used in the course to illustrate key concepts. This approach has also had quite a positive impact on our teaching of more advanced courses. Supplementary materials for this article are available online.

KEY WORDS: Pedagogy; R applet; Seminar course; Statistical literacy.

1. INTRODUCTION

The Faculty of Arts and Science at the University of Toronto offers a series of First Year Seminars, originally modeled after the Freshman Seminars available at several universities in the United States. These are one- or two-semester courses that meet once a week for two hours, and enrollment in each section is capped at 24. The courses are restricted to first-year students, and students can take only one of these seminar classes. While the course can serve as a breadth requirement, it cannot be required for any program, and prerequisites are discouraged. The course is intended to serve as a small-group learning experience for the students, and the policy is generally to base assessment on written work and class discussion, but not on tests or exams.

One of us (NR) has led a two-semester seminar course several times since the inception of these courses in 1995, and in 2008–2009 we led a seminar jointly. The title of the course is *Lies, Damned Lies and Statistics*, but a more accurate title would be *Statistics in the News*. Rather than following any text or syllabus, we generate the topics discussed each week from a haphazard selection of news items. In spite of never knowing exactly what will be covered, the pattern of topics that has emerged during each of several years' experience is not very far from the topics covered in a more traditional introductory statistics course. Building the course around news items adds a level of tension to the preparation, but also keeps the material fresh and interesting to the instructors. The student feedback for the course is generally very positive. The small-group experience is quite valued, as many of the students will be in sections of several hundred for most of their other first-year courses.

The students' backgrounds are varied, with about half of the class having exposure to mathematics in their final year of high school, and invariably some of the class being completely "math-phobic." We generally teach the course without expectations that the students have (or remember) high school exposure to senior-level mathematics or statistics. Perhaps reflecting the large urban population from which the University of Toronto draws the bulk of its students, the students are relatively independent in their work and study habits.

The format for the course was modeled on the Chance course developed at Dartmouth College by J. Laurie Snell, described by Finn and Snell (1992); see also the article by Snell (1996). The Chance course reflected, in a very original way, several pioneering ideas about statistical thinking that were emerging in the literature at the time. *Statistics: A Guide to the Unknown* (Tanur and Mosteller 1972) and *Statistics* (Freedman, Pisani, and Purves 1978) were pioneering efforts in the newly emerging science of statistical literacy. These were perhaps the first professional responses, at least in textbook length, to Huff's (1954) *How to Lie with Statistics*. In 1983, Tufte published *The Visual Display of Quantitative Information*, which by its design and choice of content highlighted the art in statistical ideas, and statistical ideas in art. The book rightly took to task many published graphical displays, and continues to be among the best resources on statistical graphics.

Many teachers of statistics in high school, college, and university now use news items and web resources as an essential part of their courses, and we make no claim to originality in this area. The works of Utts (2005) and Lohr (2009) are just two examples of textbooks that incorporate news items, and Madison et al. (2010) have compiled a casebook of news items. The work by Paulos (1995) is another example of a popular book with similar goals. Our purpose in writing this article is to describe how the course structure has developed over the years, to highlight some of the most interesting items that have emerged, and to give special emphasis to computer demonstrations we introduced in 2008–2009 to illustrate various topics. These demonstrations were all done in R (R Development Core Team 2009), and the code is available in the supplementary material.

2. THERE IS NO COURSE OUTLINE

There really is no set of prescribed topics, but some framework for choosing news items is needed. To give some level of

Jean-François Plante is Assistant Professor at HEC Montréal, Management Sciences Department, 3000 chemin de la Côte-Sainte-Catherine, Montréal QC, Canada, H3T 2A7 (E-mail: jfplante@hec.ca). Nancy Reid is University Professor of Statistics and Canada Research Chair in Statistical Theory and Applications at the University of Toronto, Department of Statistics, 100 St. George Street, Toronto ON, Canada, M5S 3G3 (E-mail: reid@utstat.utoronto.ca). For partial financial support, thanks are due to the Natural Sciences and Engineering Research Council of Canada.

  DOI: 10.1198/tast.2011.11018

comfort to the students, and the instructor, we usually start the course by reading a book. In the mid-1990s this was the book by Tufte (1983); a discussion of good and bad graphics is very accessible to students who are intimidated by quantitative thinking. More recently we have used the work of Rosenthal (2005), which has been received very positively. Other popular books which have been well-received include those by Levitt and Dubner (2005), Senn (2003), and Ayres (2007). There are new books in this vein appearing regularly; two good recent additions are those by Goldacre (2009) and Taleb (2010). Sources of news items are varied and quite haphazard, but include newspapers (Toronto's *Globe & Mail*, *The New York Times*, *The Washington Post*), magazines (*Science*, *The Economist*, *Scientific American*, *The New England Journal of Medicine*), online sources (*Wired*, *Chance News*, *Salon*), radio and television news.

Each week the students receive a handout with required reading and homework for the next week, a selection of headlines, and sometimes additional background material in the form of "technical notes." A sample handout is provided in the supplementary materials. One of the first topics covered in the course is a discussion of the distinctions among polls, surveys, observational studies, and experiments. At the same time as these topics are discussed through news items, the students are reading background material on probability (Rosenthal 2005) or graphical displays (Tufte 1983), and are asked to start finding headlines themselves related to one of these topics.

A very successful innovation, introduced by JF in 2008, was "Question of the week," in which students were encouraged to ask any question at all arising from their readings of news items or other material. We emphasized that no question was too simple, and questions ranged from the very basic, such as the difference between percentage and percentile, to the quite sophisticated. These questions often led us to study one or more news items that we had overlooked. A particularly interesting example came up early in the course: the city of Toronto hosts an all-night arts fair, *Nuit Blanche*, in early October. Activities are free, and take place at a number of venues across the downtown core. The mayor stated that one million people attend *Nuit Blanche*, and a perceptive student asked where this number came from, as there was no obvious means of counting participants, since no tickets are sold. It does seem high; the population of Toronto is 2.5 million, and the population of the Greater Toronto Area is 5.5 million. Information on this was very hard to come by, although the main web page where the number was given also referred vaguely to a survey indicating that approximately 100,000 tourists visit Toronto for Nuit Blanche, so clearly someone was trying to assess its popularity. Some rudimentary calculations of the available street area and number of venues suggested that the number might be closer to half a million. The issue of counting participation arises in several contexts: Best (2001) has a very interesting discussion of the same problem in the context of the "Million Man March" in Washington.

Within the broad topic of types of studies and surveys, we discuss confounding, sampling and non-sampling biases, and sources of information. Aspects of probability theory that arise include basic definitions of probability, as frequencies and as degrees of belief, the law of large numbers, the central limit theorem, conditional probability, and Bayes' theorem. There is very often a news item in which either the headline writer, or the journalist, confuses the probability of A, conditional on B, with the probability of B, conditional on A. As well there is nearly always an opportunity to discuss the use of Bayes' theorem in medical testing. The concepts of false positives and false negatives are very easy to convey in this context, and this comes in handy later when discussing significance testing. Rosenthal (2005) has a very nice chapter on coincidences, which is useful preparation for considerations of statistical significance later in the course.

Any number of health studies, regularly reported, lead to discussions of summaries by means and standard deviations, and their comparison using the normal distribution. Clinical trials are usually discussed in a health context as well, as is the distinction between Phase I, II, and III trials. Simple linear regression often arises in one form or another, and is initially presented as an extension of summaries by means, and later via least squares. Confounding variables, in regression and in other contexts, arise in most discussions of articles on health, nutrition, education, and various other topics. Some less common topics that arise quite often include standardized mortality ratios, life tables, aspects of the analysis of survival data and the influence of censoring, meta-analysis, relative risk, nonlinear regression, Poisson processes, time series and price indices, and ethics in statistical practice. In fact all the topics listed in the table of contents of, for example, the book by Utts (2005), have arisen at one time or another, sometimes only superficially, but sometimes considered in detail. Many of these topics do not appear in most introductory texts and courses.

The next section will illustrate by example how these topics are covered in the context of the news, by summarizing some of the more memorable or successful items covered.

3. STATISTICAL THINKING IN ACTION

3.1 "Dollar Plunges on Fears of Yes Vote"

There is invariably an interesting election in the United States or in Canada to follow during the fall semester, but without question the most exciting was the referendum for Quebec separation on October 30, 1995. The referendum asked the question "Do you agree that Quebec should become sovereign after having made a formal offer to Canada for a new economic and political partnership within the scope of the bill respecting the future of Quebec and of the agreement signed on June 12, 1995?" The title of this subsection was the headline that appeared on October 21, 1995, and Table 1 shows the series of polls taken up to that date. A "Vote no" campaign started in other parts of Canada, and thousands of Canadians descended on Montreal to hold a large rally on October 27. The referendum was defeated by the very narrow margin of 50.58% "No" to 49.42% "Yes."

While the result was arguably too close to call, an important statistical point that arose was the large percentage of those polled who reported that they were undecided, and this led to a discussion of methods that polling companies use to allocate this undecided vote. There was also some concern that

Table 1. Table of polling results in the Toronto *Global & Mail*, Wednesday, October 18, 1995 and Saturday, October 21, 1995. On October 24, 1995 this seemed to be the definitive summary of the polls, although it was not 100% consistent with earlier published stories: the Compas poll of Sept. 11–14 was earlier reported as having a sample size of 500. The Léger & Léger poll of Sept. 25–28 was earlier reported as covering the period Sept. 21–25.

Polling company	Number polled	Date	No	Yes	Undecided/ won't say
L&L	959	Sept. 7–8	42.9	43.8	13
SOM	1003	8–12	45.0	37.0	18
Compas	959	11–14	40.0	36.0	24
Createc	1004	15–19	46.2	39.8	15
Crop	2020	20–25	47	39	14
SOM	1820	19–25	48	39	13
L&L	1006	25–28	45.1	43.8	11
L&L	1015	Oct. 1–3	44	43	13
Gallup	1013	10–12	43	39	18
L&L	1002	8–12	42	45	13
SOM	981	13–16	43.4	42.9	13.6
L&L	1005	16–20	42.3	45.7	12.1

some respondents were not willing to state to Francophone interviewers that they were planning to vote "No." A very similar issue arose around early polling for the U.S. Presidential Election of 2008, where concerns were raised that the apparent support for Obama in the polls was stronger than the support would be on voting day, although these concerns turned out to be unwarranted. Students of American political history will remember vividly the election of 2000, where the popular vote for Gore and Bush was essentially tied. This was at the time quite surprising to most observers, as election results are generally thought of as the outcome of a complete count. The fact that these counts themselves have errors that can be difficult to resolve led to some very interesting statistical issues: see, for example, the works of Andrews and Feuerverger (2005) and Wolter et al. (2003).

3.2 "New Study Estimating Number of Dead in Iraq Hotly Contested"

This headline appeared in the *Global & Mail* on October 12, 2006; similar articles appeared in several newspapers, news magazines, and on wire services. The headline was based on the work of Burnham et al. (2006), which reported on the results of a cross-sectional cluster sample survey. The numbers were headline-grabbing: the Iraq body count was at the time standing at 48,693 deaths, and the estimate of Burnham et al. (2006) was 601,000 (with a margin of error of nearly 200,000), although the two sources were not measuring exactly the same quantity. There was also an exchange in *Science*: a news article was published with the headline "Iraqi death estimates called too high: methods faulted" followed by a reply from Burnham et al. defending their methods. This led to an interesting discussion of difficulties and biases that can result in sampling under difficult conditions. Two articles published in 2008 (Iraq Family Health Survey Study Group 2008; Johnson et al. 2008) gave a new estimate of 151,000 deaths and provide an excellent discussion of the possible causes of overestimation by Burnham et al. (2006).

Another headline-grabbing survey result that was patently silly, and for that reason lots of fun, appeared in the Toronto paper in October 1996: "Welfare recipients found jobs, study says" (Blackwell 1996). The conservative government that was elected in 1995 immediately cut social assistance (welfare) payments by 19%, and some months later commissioned a survey to find out why the number of people receiving social assistance had in fact declined by some 180,000 people. The article dutifully reported the results of the survey following the usual newspaper conventions: the sample size was 2100, and thus the margin of error was "plus or minus 2 percentage points, 19 times out of 20." The article also noted, in passing, that the survey was conducted by telephone. A few minutes' thought by the students raised the alarm—it is difficult to find homeless people using telephone surveys, and cuts to social assistance payments might reasonably be expected to cause an increase in homelessness.

Homelessness became an issue again in 2009, when a radio news item announced that a new survey to estimate the number of homeless in Toronto would incorporate a team of actors, who were to be paid $100 per day to mingle about and pretend to be homeless. A question raised during the radio program was whether this money might be better spent actually helping the homeless. The hapless city employee who was being badgered by the radio journalist said "We have to do it this way to make it statistically valid," and this more or less ended the conversation. They were of course using a version of capture-recapture sampling, which has been implemented for the same purpose, apparently without incident, in several major cities in the United States including New York, Boston, and Philadelphia. By a stroke of good luck, the *New York Times* had an article the same week (Dean 2009) on the use of capture–recapture sampling to estimate the size of the right whale population.

3.3 "Can Chocolate Save Your Life?"

Studies on nutrition and health are perennial headline-grabbers, and we have found the students to be more interested in these than in studies of chronic diseases. This headline, again in Toronto's *Global & Mail*, appeared on the front page of the paper on August 10, 2005, to highlight their regular column on nutrition and health. (To be fair, it should be noted that the headline writers were not the authors of the articles.) The title of the article was "Chocolate: the sweet new prescription." The article was prompted by the publication of a randomized, controlled, cross-over trial of 10 men and 10 women, for a three-week period (Grassi et al. 2005). All the study participants had elevated blood pressure, but were not taking blood pressure medication, and the two treatments to which they were randomized were 100 g daily of dark chocolate or 90 g daily of white chocolate, with further instruction to avoid other foods rich in flavonols, and guidelines on maintaining calorie intake to avoid weight gain. A number of endpoints were measured, but the headline grabber was a reduction of 11 mmHg in systolic blood pressure after 15 days of dark chocolate consumption, with no attendant change in blood pressure following treatment with white chocolate. The source article is relatively readable, and with guidance the students can extract the information from the source that appeared (or not) in the news item. This example was used to motivate more careful study of the textbook treatment of comparing means. Interestingly, the fact that each subject served as his or her own control, through the cross-over design, does not seem to have been used in the analysis.

Even to beginning students, a study based on 20 study subjects and running for three weeks seems unlikely to generalize very well, but fortunately there are many other studies bearing on the same issue. In fact a meta-analysis of studies like the one described above recently appeared (Desch et al. 2010), and as it happens the Grassi et al. (2005) study showed the largest drop in blood pressure; the average decrease in systolic blood pressure over the 10 randomized controlled trials included in the meta-analysis was a more modest 4.5 mmHg. In summer 2010 a larger meta-analysis (Ried et al. 2010) estimated a drop of 3.9 mmHg. Meta-analyses arise frequently in nutrition studies, and present a good opportunity to introduce students to the Cochrane reviews (Cochrane Collaboration 2011).

Of course if we want to know the effects in a large population, we need to study a large population. In February, 2010, results from the European Prospective Investigation into Cancer (EPIC) presented results of a ten-year prospective study of diet and health. Buijsse et al. (2010) focused on the effects of chocolate consumption on blood pressure and on the incidence of cardiovascular events. The observed decrease in systolic blood pressure between the top and bottom quartiles of chocolate consumption was just 1 mmHg. There are a number of confounding factors in this observational study, that could be controlled for in the randomized trials, and the students can quickly suggest a number of them. However, the most important difference for this issue was probably the fact that the "dose" range was only 6 g per day. Many of the confounding factors were recognized and adjusted for using regression methods, so this study could be revisited after studying simple and multiple linear regression.

Since most health studies report results in terms of relative risk, a mini-lecture on measures of absolute and relative risk typically fits in well. The EPIC study was somewhat unusual in reporting, in its press release, that the 39% observed reduction in relative risk of a heart attack or stroke could be interpreted as: "if people in the group eating the least amount of chocolate increased their chocolate intake by six grams a day, 85 fewer heart attacks and strokes per 10,000 people could be expected to occur over a period of about ten years." The website of Spiegelhalter et al. (2011) presents an entertaining and accurate discussion of the quantification of risk.

3.4 "And the Winner Is..."

The Academy Awards are a reliable source of entertainment and news. Many newspapers have an online voting and/or prediction page running between the announcement of the nominations and the awards ceremony. We have had success with Iain Pardoe's work on predicting the winner using a discrete choice model and logistic regression. By that point in the course we have covered simple and multiple linear regression and the step to the logistic transform of the response is presented as an extension of the log-transform of skewed responses such as income. The articles by Pardoe (2005, 2007) are very helpful resources for the instructor.

A related series of articles that addresses quite different problems begins with that of Redelmeier and Singh (2001), in the *Annals of Internal Medicine*. This article is written in the style of medical journal articles, and is an interesting introduction to that style as well. The work purported to show that Academy Award winners live longer: the nominees and winners were compared to actors or actresses of the same gender and age that appeared in the same movie. Discussion of this article involves introducing estimated survivor curves, which we introduce as extensions of the empirical distribution function. However, although the authors mentioned the issue of survivor bias in their article, they did not correctly adjust for this bias. The discussion by Laurie Snell in the Chance News wiki (Snell 2007) is exceptionally clear. In 2006 the *Annals of Internal Medicine* published a re-analysis (Sylvestre, Huszti, and Hanley 2006), a reply by Redelmeier and Singh (2006), and an editorial. This series of articles is interesting as well in showing that the progress of scientific research really does proceed in fits and starts, and that leading scientists may disagree about details of analysis. In spite of the criticisms of Snell and Hanley, articles claiming that other select groups have longer lives continue to appear regularly in the media. Wolkewitz et al. (2010) showed how multi-state modeling can be used to understand various types of bias in survival analysis, with special emphasis on this example. A new method of correcting for "healthy survivor" bias was presented by Han et al. (2011).

3.5 "Long-Used Drug Shows New Promise for Cancer"

A particularly overstated article with this title appeared in the *Global & Mail* on January 17, 2007 (Picard 2007). The first two paragraphs read:

> Imagine, if you will, a drug that shrinks cancer cells and can even make tumours disappear. A couple of spoonfuls a day of powder in a glass of water is all you need. There are no nasty side effects like nausea and hair loss, and no damage to internal organs such as with traditional chemotherapy. And it costs only about $2 a dose. Too good to be true? Not according to a Canadian researcher who stumbled upon the potentially new anti-cancer agent called dichloroacetate, or DCA, a drug long used to treat rare metabolic disorders. 'This is one of the most exciting results I've ever had,' said Evangelos Michelakis, an associate professor of medicine at the University of Alberta in Edmonton. 'But I can't be overenthusiastic until it works in a human being.'

In spite of Michelakis's caution in the last sentence, the tone in the rest of the article was almost irresponsibly positive. The original article, by Michelakis, Webster, and Mackey (2008), is difficult, but we used it in a guided reading assignment in which the students learned that the particular drug was tested in rats: five control animals, and eight animals in each of two treatment doses; tumor growth was followed for 5 weeks. The reading was accompanied by a technical note on randomization; see the supplementary material. The lead researcher for this article later expressed some dismay with the overly enthusiastic response: one of the biggest concerns was that patients would obtain the drug through alternate channels, making it difficult or impossible to test the drug in randomized clinical trials.

This type of overly enthusiastic response to quite preliminary work appeared in a very high-profile discussion during the fall of 2009 and winter of 2010 in media reports of a radically new and promising therapy for multiple sclerosis (Zamboni et al. 2009). Followup work is ongoing, and as expected the story is much more complex; a recent study by Doepp et al. (2010) seemed to contradict the earlier results. This issue had very widespread media coverage, with several patients electing to have the new surgery performed in other countries, in spite of pleas for caution by MS neurologists.

4. ENGAGEMENT

Classroom activities were a regular and very popular part of the course. Although they take time to set up, and class time to execute, they seem to be relatively effective in conveying key concepts, especially variability and bias. As noted by Wild, Pfannkuch, and Regan (2011), physical activities are more effective and more easily understood, and we usually started with these and moved to computer simulations or demonstrations as the second part of the activity. The computations (in R) were carried out and presented during the class. Appendix 1 provides a list of R programs whose code can be found as a zip file in the supplementary material. In this section we highlight some activities that seemed particularly successful. There are a number of resources for such activities; we found the book by Gelman and Nolan (2002) to be very helpful.

4.1 The Numbers Behind Numb3rs

The television show "Numb3rs" followed the usual police-story format, with the not-so-usual addition of a mathematician on the team of investigators. Devlin and Lorden (2007) give very clear expositions of the mathematical background, along with plot synopses, for several episodes. The format of the show was familiar to students from so many similar shows that we found they could read selected chapters of the book by Devlin and Lorden (2007) and pick up the main plot ideas very quickly.

Because Bayesian methods of analysis are often mentioned in news articles that provide some statistical detail, and because students find it interesting to learn that statisticians have had sometimes heated debates on the use of Bayesian and non-Bayesian methods, we had some success with chapter 6 of Devlin and Lorden (2007), which describes an episode in which an inmate escapes after promising to kill the main witness of his trial. Charlie, the mathematician, uses sighting reports in a Bayesian analysis to determine the places where the inmate is likely to hide. This enables Don, the FBI agent, to correctly guess the location of the inmate, who naturally is caught just in time to prevent the murder.

Our program `Bayes.R` first illustrates the use of Bayes' theorem for estimating the probability of heads under simple coin tossing, using a uniform prior. We let the students flip a selected coin and plot the updated posterior as we collect more data, illustrating that it converges to $p = 0.5$. The second part of the program uses `image` plots to show an area of interest: sighting reports are clicked on the plot and the probable locations of the escapee are updated, illustrating the process described by Devlin and Lorden (2007).

Two other news articles that highlight the role, and potential controversy, of Bayesian analysis are a discussion of climate models reported in *Science* (Stainforth et al. 2007) and an article on cognitive reasoning in the *Economist* (Griffiths and Tenenbaum 2006).

4.2 Capture–Recapture Sampling

As mentioned in Section 3.2, capture–recapture sampling arose in the news in two quite different contexts, within the same week. One was the more conventional use of this in animal abundance and the other in human sampling. We used the work of Hammond (2006) as an accessible reference for the students. The activity that we used to illustrate this involved a very large bag of small, wrapped, chocolates. We transferred these chocolates to a larger container, so that they could be more easily shuffled and sampled, and then we passed this around the room, to students working in pairs. The first pair chose 10 chocolates, marked the wrappers with a black permanent marker, and returned them to the box. The second pair chose 10 more, and called out how many of their sample were marked: this provided the first estimate of the total number. This pair then marked their 10 with a red permanent marker, returned them to the box, and so on. Based on each subsequent sample, we used the data on the number and color to construct estimates of the population total, along with estimated confidence intervals, and updated these on the screen. One of the most striking things about the exercise was the imprecision of the estimates. It was also instructive for the students to see that the standard calculations do assume a model analogous to our bucket of chocolates; they could quickly see that more realistic modeling for human and animal populations would be a natural next step. There is of course a large literature on exactly that; see for example, the articles by Rivest (2008) and Baillargeon and Rivest (2007).

4.3 Polling M&Ms

With every election or referendum, results of numerous polls are published in the news, typically with confidence intervals. In class, a jumbo-sized bag of M&Ms can be used to represent the voters. Their colors can easily be paired to the different candidates, or in Canada, political parties. Every student in class can then become a pollster and produce their own predictions for the "elections," in the form of a confidence interval. We used an R applet to plot the intervals obtained by each student and to compare them. Figure 1 displays the results we obtained in the Fall of 2008. Seeing about 95% of the class obtain intervals that include the true proportion helps in understanding the coverage of a confidence interval. A few years ago, the true proportions of every color of M&M was given online, but this information is no longer available. An alternative is to consider the whole bag as the population of interest and make sure that all M&Ms are observed.

In a business setting, colors of the M&Ms may also be mapped to customers' level of satisfaction, or to any other characteristic expressed by a categorical variable.

4.4 The Stock Market Crash

There was more than usual interest in finance and the stock market in 2008–2009. The programs in `stock.R` were used

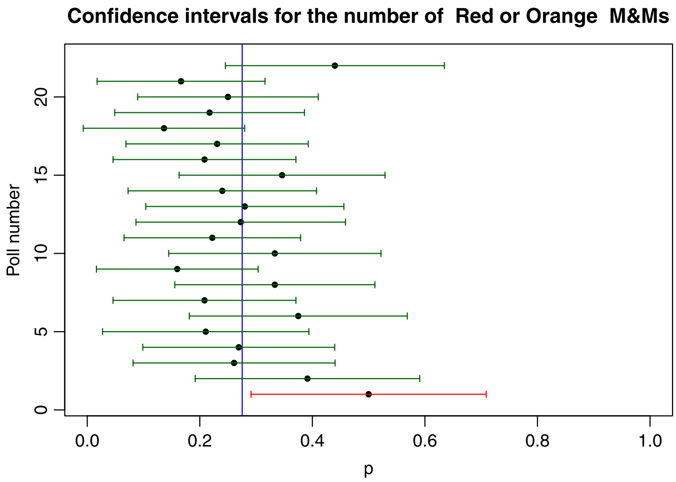

Figure 1. Confidence intervals obtained by the students for the M&Ms sampling experiment. The online version of this figure is in color.

to aid in grading a homework exercise that was quite popular. (The homework is provided in the supplementary materials along with a sample solution.) Each student was instructed to choose a stock of interest, to obtain from *finance.yahoo.com* a 14-year record of its performance, and to download this into a spreadsheet such as Excel. The students worked independently, so there was little overlap in their choice of stocks. When this homework was returned, we listed on the blackboard the value of $1000 invested in 1995, for each student. It was quite striking that as of April 2009, only Apple had showed a net gain.

Widespread underestimation of the risk of various financial products seems to be one of the main culprits in the financial crisis. An article in *Wired* (Salmon 2009) pointed at one formula "that killed Wall Street" in particular, and mainstream media quickly picked this up. As a result one of our most interesting "Questions of the Week" was to comment on the statistical formula in the *Toronto Star*. We were able to find the news item on the *Star*'s website during class (Kelly 2009). Although we had not yet heard of the *Wired* article, the formula was clearly a Gaussian copula and the article mentioned problems with the estimation of the probability of joint default. We improvised a short lecture on correlation, ranks, copulas, tail dependence, and extreme values. It was a very good opportunity to discuss the fact that all models are based on assumptions that one should not overlook. It also gave us an opportunity to mention current research in passing, and gave an excellent illustration of a few of the many things that remain to be discovered in statistics. Whenever we could link statistical concepts to a question that came from the students it seemed to generate more interest and more discussion.

4.5 Statistical Learning and Music

A student from the class watched a television show called *The Musical Brain*. When a company specializing in predicting the success of a song was featured, she wondered what statistical methods might hide behind the story. A radio interview on the same topic was aired by National Public Radio (2007). While the exact methodology was obviously kept secret, data mining tools were clearly used. We took this opportunity to discuss data mining in general and to present support vector machines in more detail. We found a freely available dataset by Homburg et al. (2005) that contained characteristics of different songs, automatically extracted from mp3's. These characteristics can be used to predict the type of music (rock, hip hop, etc.) with surprisingly good results. Students seemed impressed that in a matter of hours, we could illustrate predictions similar to those in the television show.

5. CONCLUSION

The course described here is quite specialized, and since the group is small, it is relatively easy to administer. Most statistics courses and statistics instructors have many more constraints. We have found, though, that the use of articles from the news is quite addictive for the instructor, and transfers easily to a number of other courses, including design of experiments, methods of applied statistics, sample surveys, data mining and machine learning, probability, multivariate analysis, and even advanced courses on statistical inference.

Although it was not obvious to us when we started, there was no difficulty in finding enough news each week. On the contrary, it was sometimes difficult to keep the number of articles to a reasonable limit, especially as one gets more addicted to the process. Another surprise, now familiar, is that a large fraction of freshman do not follow current events, neither reading the newspaper nor reading online sources of news. However, we view exposure to current events as part of their introduction to university, no matter what their proposed choice of major.

While it varies somewhat with the class, there are some types of articles that fall quite flat, and others that ring in very well. Topics that do not seem to work very well include chronic diseases (with the exception of highly profiled "miracle" cures for cancer), parenting, gender equity, physical science, and population health. We have also been surprised at the lack of interest in sports and gaming and in economics and finance. Topics that are generally well-received include diet and health, education, law and forensics, public policy, and issues surrounding the environment. This would undoubtedly vary considerably from one university environment to another, and indeed a reader of an earlier draft stated "I can't imagine that sports doesn't work well in the U.S." One strategy we have used is to have the class vote on a collection of substantive areas about one-third of the way through the course; this determines selection of articles for the remainder of the course.

There is a danger of over-skepticism that we continue to grapple with. Many newspaper articles, and especially their headlines, are overstated, and after reviewing several dozen of these and pointing out the inevitable caveats, a rather gloomy picture tends to emerge. This was brought home to us by one plaintive question near the end of the course: "Can we believe any of these studies?" To balance this we try to find and comment positively on work that seems to be of very high quality. Of course the main point is that knowledge is not usually advanced by a single study, and the interesting problems surround how the picture slowly unfolds.

Although we do not have a syllabus, and have the luxury of choosing material to cover, in recent years the students indicated that they would like to have a textbook for background reading. We have found the book by Utts (2005) very well suited for this purpose; although we do not go through the book at all systematically, many of the issues that come up are addressed there, and the more mathematically able students enjoy having some access to formulas and a more structured approach to the subject.

As with most teaching, the rewards come in surprising moments and throw-away comments: the student who missed his subway stop because he was engrossed in *Struck by Lightning* (Rosenthal 2005); the philosophy student who decided to add a science major; the math-phobe who figured out that numbers were neither incomprehensible nor mysterious.

APPENDIX: R APPLETS FOR ACTIVITIES AND DEMONSTRATIONS IN CLASS

In the 2008–2009 version of the course, we regularly used applets to illustrate statistical concepts or to perform calculations based on the data collected in a class activity. The programs described below can be found in the archive Rapplet.zip provided in the supplementary materials.

- `Bayes.R`: The first section of this script is used for a class activity. We choose a coin and pass it around. Students are asked to flip it multiple times. The results are used to update the distribution of the probability of getting a head. We start with a uniform prior and as the number of flips increases, the posterior takes shape around a probability of $1/2$.

  The second part of this script is inspired from chapter 6 of Devlin and Lorden (2007) that describes how a Bayesian analysis can be used to zoom in on the location of a fugitive based on sighting reports. The demo starts with a uniform distribution over the city (a square) and the user can click locations of the sighting. The posterior is updated and shows the probable location of the criminal using a color-coded map.
- `CLT-coins.R`: To illustrate the Central Limit Theorem in action, we brought a full cup of pennies to class. We passed pennies around and asked students to record how many heads they had on 5, 10, and 20 flips. The script traces the histograms for the number of heads obtained by the students and simulates the same experiment with more flips and more students.
- `CLT-truedist.R`: To illustrate further the Central Limit Theorem, this script draws the exact distribution of a sum of independent variables. Three variables are considered: a Bernoulli (representing a coin flip), a uniform on $\{1, 2, \ldots, 6\}$ (representing the toss of a die), and an exponential variable (that can be associated with waiting times).
- `MMs.R`: This script is used for a class activity on confidence intervals. A jumbo-sized bag of M&Ms is brought to class. In the context of an election, the colors are associated with political parties, or candidates, and the students become pollsters. In a business setting, the colors can also represent the satisfaction level of customers, or their segment. After drawing a sample, every student is asked to calculate a 95% confidence interval for the proportion of voters for a given party. Data are collected and the script draws all the intervals on a single plot with the true proportion. Such a plot shows clearly what a significance level means: on average, 95% of polls produce an interval containing the true value, but no single poll can make a statement on probability on his own success once the sample is selected.
- `capture.R`: A short script shows on a plot the concept of the line transect method for estimating an animal population. Another applet shows the idea of capture-recapture methods. Finally, the `Rcapture` package is used to estimate the number of chocolates in the box for the class activity on capture-recapture described in Section 4.2.
- `diffmeans.R`: This script is used to illustrate the distribution of the difference of two random variables (height or weight of males versus females) in the context of hypothesis testing.
- `montecarlo.R`: As examples of Monte Carlo experiment, we illustrate numerical integration and the Buffon needle problem.
- `regression.R`: A series of classical examples are first presented to illustrate different uses of a regression. To understand the fit, an interactive script lets the user choose the "best line" interactively by eye. The best solution so far is kept and its sum of squares is returned to be compared to the optimal solution.
- `stock.R`: This script returns a personalized pdf solution to the final assignment for any given stock. LaTeX is required.
- `svm.R`: Different graphical examples are shown to illustrate how support vector machines classify bivariate data. We also use the dataset of Homburg et al. (2005) to predict the type of music based on computer extracted features of mp3 files.

SUPPLEMENTARY MATERIALS

Supplement: *R applets*: The code of the applets presented in Appendix 1 (zip file). *Handout*: A sample handout from the course (pdf file). *Homework*: The questions of the final assignment (pdf file). *Solution of the homework*: A sample of the personalized solutions we prepared for the students using the code in stock.R (pdf file). *Techrandom*: A typical "technical note"; this one on randomization. All files are contained in a single archive (and can be obtained via single download). (SuppMat.zip)

[Received February 2011. Revised May 2011.]

REFERENCES

Andrews, D. A., and Feuerverger, A. (2005), "The Statistical Significance of Palm Beach County," in *Statistical Modelling and Analysis for Complex Data Problems*, eds. P. Duchesne and B. Rémillard, New York: Springer-Verlag, pp. 17–40. [82]

Ayres, I. (2007), *Supercrunchers*, New York: Random House. [81]

Baillargeon, S., and Rivest, L.-P. (2007), "Rcapture: Loglinear Models for Capture–Recapture in R," *Journal of Statistcal Software*, 19 (5), available at *http://www.jstatsoft.org/v19/i05*. [84]

Best, J. (2001), *Damned Lies and Statistics*, Berkeley: University of California Press. [81]

Blackwell, T. (1996), "Welfare Recipients Found Jobs, Study Says," *Toronto Star*, 24 October, 1996. [82]

Buijsse, B., Weikert, C., Drogan, D., Bergmann, M., and Boeing, H. (2010), "Chocolate Consumption in Relation to Blood Pressure and Risk of Cardiovascular Disease in German Adults," *European Heart Journal*, 31, 1616–1623. [83]

Burnham, G. M., Lafta, R., Doocy, S., and Roberts, L. (2006), "Mortality After the 2003 Invasion of Iraq: A Cross-Sectional Cluster Sample Survey," *The Lancet*, 368, 1421–1428. [82]

Cochrane Collaboration (2011), "What Are Cochrane Reviews?" available at *http://www.cochrane.org/cochrane-reviews*. [83]

Dean, C. (2009), "The Fall and Rise of the Right Whale," *The New York Times*, March 16, available at *http://www.nytimes.com/2009/03/17/science/17whal.html?_r=1*. [82]

Desch, S., Schmidt, J., Kobler, D., Sonnabend, M., Eitel, I., Sareban, M., Rahimi, K., Schuler, G., and Thiele, H. (2010), "Effect of Cocoa Products on Blood Pressure: Systematic Review and Meta-Analysis," *American Journal of Hypertension*, 23, 97–103. [83]

Devlin, K., and Lorden, G. (2007), *The Numbers Behind NUMB3RS: Solving Crime With Mathematics*, New York: Plume. [84,86]

Doepp, F., Friedemann, P., Valdueza, J. M., Schmierer, K., and Schreiber, S. J. (2010), "No Cerebro-Cervical Venous Congestion in Patients With Multiple Sclerosis," *Annals of Neurology*, published online June 11, 2010. DOI: *10.1002/ana.22085*, available at *http://onlinelibrary.wiley.com/doi/10.1002/ana.22085/abstract*. [84]

Finn, J., and Snell, J. L. (1992), "A Course Called 'Chance'," *Chance*, 5, 12–16. [80]

Freedman, D., Pisani, R., and Purves, R. (1978), *Statistics*, New York: W. W. Norton. [80]

Gelman, A., and Nolan, D. (2002), *Teaching Statistics: A Bag of Tricks*, Oxford: Oxford University Press. [84]

Goldacre, B. (2009), *Bad Science*, London: Harper Perennial. [81]

Grassi, D., Lippi, C., Necozione, S., Desideri, G., and Ferri, C. (2005), "Short-Term Administration of Dark Chocolate Is Followed by a Significant Increase in Insulin Sensitivity and a Decrease in Blood Pressure in Healthy Persons," *The American Journal of Clinical Nutrition*, 81, 611–614. [82,83]

Griffiths, T. L., and Tenenbaum, J. B. (2006), "Optimal Predictions in Everyday Cognition," *Psychological Science*, 17, 767–773. Discussed in *The Economist*, Jan. 5, 2006, available at *http://www.economist.com/node/5354696*. [84]

Hammond, P. (2006), "Whale Science—And How (Not) to Use It," *Significance Magazine*, 3, 54–58. [84]

Han, X., Small, D., Foster, D., and Patel, V. (2011), "The Effect of Winning an Oscar Award on Survival: Correcting for Healthy Performer Survivor Bias With a Rank Preserving Structural Accelerated Failure Time Model," *The Annals of Applied Statistics*, to appear. [83]

Homburg, H., Mierswa, I., Möller, B., Morik, K., and Wuest, M. (2005), "A Benchmark Dataset for Audio Classification and Clustering," in *Proceedings of the International Symposium on Music Information Retrieval 2005*, eds. J. D. Reiss and G. A. Wiggins, London: Queen Mary University, pp. 528–531. [85,86]

Huff, D. (1954), *How to Lie With Statistics*, London: W. W. Norton. [80]

Iraq Family Health Survey Study Group (2008), "Violence-Related Mortality in Iraq From 2002 to 2006," *New England Journal of Medicine*, 358, 484–493. See also Letters to the Editor, *New England Journal of Medicine*, 359, 431–434. [82]

Johnson, N. F., Spagat, M., Gourley, S., Onnela, J.-P., and Reinert, G. (2008), "Bias in Epidemiological Studies of Conflict Mortality," *Journal of Peace Research*, 45, 653–663. [82]

Kelly, C. (2009), "Meet the Man Whose Big Idea Felled Wall Street," *Toronto Star*, March 18, available at *http://www.thestar.com/Business/article/604033*. [85]

Levitt, S. D., and Dubner, S. J. (2005), *Freakonomics*, New York: Harper Collins. [81]

Lohr, S. L. (2009), *Sampling: Design and Analysis*, North Scituate: Duxbury Press. [80]

Madison, B. L., Boersma, S., Diefenderfer, C. L., and Dingman, S. W. (2010), *Case Studies for Quantitative Reasoning: A Casebook of Media Articles* (2nd ed.), New York: Pearson. [80]

Michelakis, E. D., Webster, L., and Mackey, J. R. (2008), "Dichloroacetate (DCA) as a Potential Metabolic-Targeting Therapy for Cancer," *British Journal of Cancer*, 99, 989–994. [83]

National Public Radio (2007), "Predicting Popularity: The Math Behind Hit Music," May 25, available at *http://www.npr.org/templates/story/story.php?storyId=10442377*. [85]

Pardoe, I. (2005), "Just How Predictable Are the Oscars?" *Chance*, 18, 32–39. [83]

——— (2007), "Predicting Oscar Winners," *Significance*, 4, 168–173. [83]

Paulos, J. A. (1995), *A Mathematician Reads the Newspaper*, New York: Basic Books. [80]

Picard, A. (2007), "Long-Used Drug Shows New Promise for Cancer," *The Globe & Mail*, January 17, available at *http://www.theglobeandmail.com/life/article736333.ece*. [83]

R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, available at *http://www.R-project.org*. [80]

Redelmeier, D. A., and Singh, S. M. (2001), "Survival in Academy Award-Winning Actors and Actresses," *Annals of Internal Medicine*, 134, 955–962. [83]

——— (2006), "Reanalysis of Survival of Oscar Winners," *Annals of Internal Medicine*, 145, 392. [83]

Ried, K., Sullivan, T., Fakler, P., Frank, O. R., and Stocks, N. P. (2010), "Does Chocolate Reduce Blood Pressure? A Meta-Analysis," *BMC Medicine*, 8, 39–50. Available at *http://www.biomedcentral.com/1741-7015/8/39*. [83]

Rivest, L.-P. (2008), "Why a Time Effect Often Has a Limited Impact on Capture–Recapture Estimates in Closed Populations," *Canadian Journal of Statistics*, 36, 75–84. [84]

Rosenthal, J. (2005), *Struck by Lightning: The Curious World of Probabilities*, Toronto: Harper Collins Canada. [81,86]

Salmon, F. (2009), "Recipe for Disaster: The Formula That Killed Wall Street," *Wired*, 17, 03, available at *http://www.wired.com/techbiz/it/magazine/17-03/wp_quant*. [85]

Senn, S. (2003), *Dicing With Death: Chance, Risk and Health*, Cambridge: Cambridge University Press. [81]

Snell, J. L. (1996), "Teaching a Chance Course," *Communications in Statistics—Theory and Methods*, 25, 2853–2862. [80]

——— (2007), "Oscar Winners Do Not Live Longer," available at *http://www.causeweb.org/wiki/chance/index.php/Oscar_winners_do_not_live_longer*. [83]

Spiegelhalter, D., Pearson, M., Smith, O., Garay-Arevelo, A., and Short, I. (2011), "Understanding Uncertainty," available at *http://understandinguncertainty.org/*. [83]

Stainforth, D. A., Allen, M. R., Tredger, E. R., and Smith, L. A. (2007), "Confidence, Uncertainty and Decision-Support Relevance in Climate Predictions," *Philosophical Transactions of the Royal Society A*, 365, 2145–2162. Discussed in *The Economist*, Aug. 18, 2007. Available at *http://www.economist.com/node/9645336*. [84]

Sylvestre, M.-P., Huszti, E., and Hanley, J. A. (2006), "Do Oscar Winners Live Longer Than Less Successful Peers? A Reanalysis of the Evidence," *Annals of Internal Medicine*, 145, 361–363. [83]

Taleb, N. N. (2010), *The Black Swan* (2nd ed.), New York: Random House. [81]

Tanur, J. M., and Mosteller, F. (eds.) (1972), *Statistics: A Guide to the Unknown*, San Francisco: Holden-Day. [80]

Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Cheshire: Graphics Press. [81]

Utts, J. (2005), *Seeing Through Statistics* (3rd ed.), Belmont, CA: Thompson Brooks/Cole. [80,81,86]

Wild, C. J., Pfannkuch, M., and Regan, M. (2011), "Towards More Accessible Concepts of Statistical Inference," *Journal of the Royal Statistical Society, Ser. A*, 174, 247–295. [84]

Wolkewitz, M., Allignol, A., Schumacher, M., and Beyersmann, J. (2010), "Two Pitfalls in Survival Analyses of Time-Dependent Exposure: A Case Study in a Cohort of Oscar Nominees," *The American Statististician*, 64, 205–211. [83]

Wolter, K., Jergovic, D., Moore, W., Murphy, J., and O'Muircheartaigh, C. (2003), "Reliability of the Uncertified Ballots in the 2000 Presidential Election in Florida," *The American Statistician*, 57, 1–14. [82]

Zamboni, P., Galeotti, R., Menegatti, E., Malagoni, A. M., Tacconi, G., Dall'Ara, S., Bartolomei, I., and Salvi, F. (2009), "Chronic Cerebrospinal Venous Insufficiency in Patients With Multiple Sclerosis," *Journal of Neurology, Neurosurgery & Psychiatry*, 80, 392–399. [84]