DOES STATISTICAL EDUCATION INCLUDE SOCIAL ISSUES?

Jacky Galpin
University of the Witwatersrand
South Africa

As educators, we should not only aim to provide our students with technical skills, but should also help them develop life skills. In recent times there has been an increasing emphasis on communication skills, application skills and reporting skills, but we possibly have not yet sufficiently articulated the social issues associated with good data collection, analysis and reporting. We also need to demonstrate to the students, and through them to the community, the wide field of applicability of statistical techniques, and the need for viewing events from a numerate point of view (among others), in order to interpret what the events mean. There are many social issues that can and should be raised with our students, which can also be used to illustrate statistical techniques. Examples of this, particularly pertinent in South Africa, are issues such as HIV/AIDS, rights of women, etc. For example, HIV/AIDS can be used to discuss regression on indicator variables (HIV negative, HIV+, then later expand to symptomatic and non-symptomatic). This could then be combined with a few questions about whether the class thinks that mortality tables apply to them. This paper focuses on the questions: do statisticians have a social responsibility to students to include such issues among the technical issues, and what is the best way of doing this?

INTRODUCTION

Teaching statistics at university level has a number of aspects. For main stream students, this includes teaching of theory, ability to derive new theory and ability to apply the theory. The importance of the last component has been attracting attention over the past few years, as it is no longer possible to say that this can be learnt 'on the job', if we want our students to be marketable. Comments from employers of our students have also lead to a greater emphasis on students working in groups and not just as individuals, gaining some competence in the use of statistical packages as well as word processing and spreadsheets, and in report writing. These are all 'good things' aimed at producing competent, marketable graduates. For the applied students, statistical teaching involves understanding of when to apply what technique, and how to discover problems in the analyses.

However, I would argue that one also needs to include in the students' education, aspects of social issues, and some development of social responsibility, in order to enable them to be socially responsible citizens, and not just 'holders of a degree'. I argue that statistics courses are an appropriate vehicle for at least some aspects of this. One possibility is the use of real life, or topical examples in teaching. Another useful method is project work - we need to get students used to real-life data, with all its warts (such as typos, changes over the period of data collection etc.), as well as teaching them the theory. It is sometimes possible to include socially relevant issues in such projects. In this paper, I outline my reasons for placing emphasis on social issues, as well as some methods that I have used for this. This is intended to be a paper sparking discussion.

REASONS FOR EMPHASIZING SOCIAL RESPONSIBILITY

Living in South Africa, we have a number of urgent issues to deal with, not the least of which is the HIV/AIDS epidemic. The December 2000 AIDS epidemic report of the WHO estimated the adult prevalence rate (adult=15-49 years old) for Sub-Saharan Africa, at 8.8%. A report on March 22 2001 (MSNBS News) indicates a figure of 1 in 9 or 11.1% for South African adults. The 2001 Department of Health reports based on Antenatal Clinic data is that over 25% of pregnant women are infected, with provincial rates differing, the maximum estimate being Kwa-Zulu Natal at 36.2%.

The HIV/AIDS epidemic has lead to prevalence of a number of other diseases increasing, such as TB. The common reaction of 'I feel better now, I'll stop taking the medicine' has lead to a

rise in multiple drug resistant TB. HIV positive patients are also more susceptible to various diseases (including malaria), and when they do become ill are often more seriously ill.

This is an important issue for two main reasons: our students are increasingly exposed to these risks themselves. While not completely agreeing with the statement made in the American TV program Evening Shades screened some time back 'he's only a teenager – nothing above the waist works yet', it is undeniable that our students are likely to be sexually active. We all spend a lot of effort teaching the students – I would prefer it if they did not contract AIDS or TB or such-like, and die within a few years!

In addition, the students need to be educated in these matters, so as to be able to pass the message on to their peers and siblings, their parents, and their communities. Many of our students come from rural backgrounds, and can thus pass on information to hard to reach communities (although they may of course be classified as 'wet behind the ears', and thus disregarded by their elders). The students will also increasingly come up against these problems in their workplace, particularly the Actuarial Science students. Apart from this, it is predicted that the HIV/AIDS epidemic will result in the cost of belonging to a medical aid scheme will be higher than the bond on a house, within the next 3 years.

In addition we have a high death toll on the roads, increasing use of drugs, a need for greater awareness of gender issues, etc. At the time of writing, the nation is watching horrified at reports on the trial of six adult men accused of raping a 9 month old baby girl.

Students need to be made aware of what is happening with these issues, and how to deal with what is reported in the press. For example, the press continuously reports on the horrendously high death toll on the roads over the Easter and Christmas periods, but does not note that the death toll at other times of the year is about the same, it just happens in urban areas, and involves mainly pedestrians, rather than drivers of cars. Similarly, the rape of women, and the abuse of children appears to be greatly increasing over the last few years, but this may well be because society is becoming aware that this is socially unacceptable, and not the fault of the women or children involved.

HOW TO BRING SOCIAL ISSUES INTO TEACHING?

Teaching of statistics can easily be used to include social issues, since we are teaching statistical literacy as well as the nuts and bolts of our profession. The question is just how best to do this. I mainly include these issues in the examples used in class (with anecdotes), and in exam questions, and digressions in class re current items in the press, TV and radio. It is not clear how much the attention the students pay to the actual example in an exam situation, but I think it may be useful as a reinforcement (and can't do any harm).

The biggest problem with all of these is finding appropriate data. It is hard to obtain South African medical data in many areas (other than via reports in the press, or in collaborative work with medical personnel). Two examples of inclusion in teaching material are:

a) While teaching indicator variable regression: get the students to set up the indicators for HIV positive and HIV negative. Then get them to extend this to 3 groups: HIV positive symptomatic, HIV positive non-symptomatic, and HIV negative.

   For the main stream mathematical statistics students, where I teach regression to the 3rd year class, I then digress into a discussion of whether mortality tables apply to the students or not. Most have done mortality tables during their first year, all believe the tables are correct, but no-one believes they apply to them! I then ask whether they know how to guard against getting AIDS by practicing safe sex, and suggest to the class that they get further information from the Campus Health Clinic. With 3rd year students (average age 21), this does result in some embarrassment.

   For the applied students, where I teach a class from a number of different faculties at honours (=4th year) level, it is easier, as the students are just that little bit older, and have not all been together for a number of years. Again, I use the example of indicator variables in regression, but often also include this in contingency table analysis etc. It is, however, sometimes difficult to get this across to finance students.

b) While teaching contingency tables and log-linear model, one can bring in such examples as:

- the increased risk of coronary disease with smoking; here I often bring in anecdotes - I was lecturing a bunch of medicos, and one came in fuming: Stupid woman says she is now going to give up smoking, I told her 'you've had your heart attack, it's too late now'. Also a friend recently died of a stroke, due to years of smoking and messing up his arteries.
- the increased risk of dying or being severely injured in a car accident, if not wearing a seat belt; (a feature in the news in SA during the Easter and Christmas periods).
- the differing risks of a baby contracting AIDS depending on the type of birth (normal or caesarian, duration of labour of the mother, and whether the mother breastfed or not.

In my classes I am also trying to teach students how to draw together material from different courses. To this end I sometimes digress from the pure regression or whatever material, into a discussion of newspaper/TV items of interest (and then relate the principle back to the material I am teaching). Some examples are:

a) Rape is increasing – well… that actually means reported rapes. Do we throw a party and say it's because of the increased rate of reporting for the same number of rapes? Or go into sackcloth and ashes and say it's the same rate of reporting for an increased number of rapes? Or a combination of the two? There has been a campaign in South Africa to remove the stigma of reporting rape. (15 years ago, reporting a rape meant a woman was stigmatized as being promiscuous.) Can this be used to bring out the point of it not being possibly to imply causation from statistics in many cases? Also: what can you use the data for and what can't you? Is all of the historically collected data relevant to the regression model you are trying to set up?

b) HIV/AIDS rates: Estimated from antenatal clinic figures, from public hospitals. Is there a difference in infection rates according to income? Wealthier people will go to private hospitals, where there are no long queues. Public hospitals are mainly frequented by poorer people, which in South Africa generally means a higher percentage of black people than in the general population. Is there a racial difference in HIV/AIDS incidence? Could this be biasing results? So you have the HIV rate for pregnant women – how do you get to the HIV rate for the general sexually active population? Extend to a discussion of modelling, and the assumptions you need to make. Extend to a discussion of sampling implications, and whether 'collected' data (as in regression) as opposed to designed data (as in a probability survey or designed experiment) is likely to be representative, and how one should try to check this..

I always make my students apply the techniques I have taught them to real life data (which they need to find for themselves). Making them find the data for themselves often means that they are keen to find out what the data says. It would be nice to get the students to do projects on data from these social issue areas, but often they are not interested in these areas. A colleague does do this – in teaching sampling, he gets the students to draw up a very simple questionnaire, get a sample, generally of people, ask them what they think about the issue, and then analyze the data.

A major problem with HIV/AIDS and some other areas is a lack of data. There seem to be very few Internet sites with full data sets, and where there is data, there are lots of missing values. Another problem is that some of the projects are done in groups (depending on the year of study, and size of the class). In these cases, the data chosen is often that of interest to only one of the group (or is often just anything they could find that they think will satisfy the requirements of the project). By the time one has taught the students enough that they can do a sensible analysis, exams are looming, and my students are keen to get the project out of the way, and concentrate on getting high marks.

CONCLUDING COMMENTS

Should we as teachers of statistics be raising such issues with our students? I whole-heartedly believe so, as we need to develop the students, as people not just as statisticians. Statistics is also a subject in which the more one knows about almost anything, the better one can

handle the data that comes your way. There are too many students being produced (in South Africa anyway) that are very good at the theory, but cannot make sense of real life data.

Is there a way to evaluate the methods used, given that the issues are generally brought in as digressions to the main class materials? The only area where this seems possible, is to ask the students in classes where sampling is in the syllabus, to design a sample to estimate HIV infection rates (or number of rapes, etc). But this is very embarrassing to young adults. This embarrassment could lead them to fake the data, and could lead to resentment. However, if they actually did such a project, one could see whether the students have understood the issues. Understanding of issues is of course a very different matter to accepting that one needs to take these things into account in one's own life. Can the effectiveness of these ideas be evaluated? Is there a better way, or additional ways to accomplish that? Well, the purpose of the paper is to stimulate discussion on the topic!

ACKNOWLEDGEMENT

REFERENCES
Aids Epidemic Update: December 2000. Geneva: UNAIDS and WHO.
Department of Health (South Africa) (2001). *National HIV and Syphilis Sero-prevalence Survey of Women Attending Public Antenatal Clinics in South Africa, 2000.*