PROMOTING STATISTICS THINKING AMONGST SECONDARY SCHOOL STUDENTS IN THE NATIONAL CONTEXT

Philip J. Boland
National University of Ireland
Ireland

Although secondary school students in many countries (like Ireland) get a limited and basic introduction to Statistics, it is often of a mechanical and tedious nature with little or no emphasis on data analysis and practical examples. In particular they, together with their teachers, rarely see the applicability and challenging nature of statistical thinking. Statisticians need to promote these aspects of statistics to the young, their teachers and the public at large. It is suggested that the use of examples of a local (often of a national) nature should be encouraged in an effort to emphasize the relevance of statistical thinking. With this in mind, several examples which have been successfully used in the Irish context are discussed.

INTRODUCTION

In many countries, secondary school students have a very limited vision of both the importance and impact of probability and statistics in everyday life. Their formal statistical education (for example in countries like Ireland) is often of a formulae based approach, and is rarely practically oriented (see Boland & Nicholson, 1996). In such situations there is a dire need to inform both students and their teachers about the challenging and applicable nature of statistics, as well as how statistical thinking comes into diverse areas of modern life. It is suggested that a key issue in this context is the local relevance of statistical examples, and this can be particularly important in countries that rely on others for much of their educational material.

As part of an overall policy of promoting Science in general in Ireland, presentations have been made to secondary school students, their teachers and the public highlighting the many uses of statistics and the need for more statistical thinking (see Boland, 1998). In doing so, an emphasis has been laid on trying to use examples of local or national (in this case Irish) interest. My belief is that statisticians in similar situations should endeavour to co-operate in assembling relevant material, which can be used to promote statistical thinking.

With this in mind, I intend to demonstrate this tenet through a series of examples generated for Irish secondary school students. I initially trace the historical origins of probability in games of chance, but quickly move on to highlight how statistical thinking comes into diverse areas of modern life while trying to emphasise relevance. The use of statistics in Forensic science is illustrated through an example of probabilistic/statistical evidence in the Irish trial of the murder of Lord Louis Mountbatten. The power of good graphical techniques is used to illustrate the value and/or impact of evidence in a case where two Irish University students were accused of cheating in their final exams. Other graphical techniques are also used to highlight the interesting nature of how the number of daily births in Ireland (and I expect most countries) varies by day of the week. The statistical properties of a good diagnostic test are illustrated with reference to the Tuberculin test used to detect Bovine Tuberculosis in Ireland. Incidence maps of bovine TB and human cancer in Ireland are presented and discussed, emphasizing the need for standardization in order to make relevant geographical comparisons. These and other examples are meant to highlight how we can try to promote the power and use of good statistics.

HISTORICAL ORIGINS OF PROBABILITY AND STATISTICS

Although my prime motivation in addressing students and their teachers is to instil ideas about statistical thinking and the modern problems it can address, there is still a certain fascination about probability and gambling. I have found, for instance, that adding the word LOTTO to the title of a presentation on the practice of statistics will almost certainly lead to a larger audience (see Boland & Pawitan, 1999). Therefore I often begin with a few words about the historical origins of probability and gambling, followed by some comments on the Irish National Lottery game LOTTO.

- The foundations of probability theory lie in problems associated with gambling and games of chance, and interesting developments came from correspondence between the philosopher Blaise Pascal and the mathematician Pierre de Fermat (1654) relating to what is sometimes called The Problem of Points. In a simplified version of the problem, I ask students to imagine that two players A and B contribute £6 each to a pool (total of £12). A game is to be played whereby a fair coin is to be tossed until 3 heads occurs (whereby A wins) or 3 tails occurs (B wins). Unfortunately the game is unexpectedly stopped after only 3 flips with the result: {H, H, T}. The question is how, in this case, should the pool of £12 be split? I try to get the audience to participate and put forth various arguments for how the pool might be appropriately dispersed. Two popular arguments which usually leads to an interesting discussion involving the relative merits of each are:Player A has won 2/3 of the games, and therefore should take 2/3 of £12 = £8.
    - If play were to continue, B would win if the next two games yield TT (with probability 1/4), and otherwise A wins (with probability 3/4). Hence A should take 3/4 of £12 =£9. The Irish National Lottery successfully runs the LOTTO 6/42 game twice weekly, and guarantees a minimum jackpot of one million punts (approximately 1,250,000 Euro). Although the game has the advantage that for a given amount of money everyone has the same chances of winning a share in the jackpot, only half of the money put into the game by the participants is returned to the players. Moreover, the odds of winning a share of the jackpot are extremely small. In order to give some indication of the odds of winning various prizes, I discuss the average time it takes to win them if one plays the minimum amount twice weekly (which enables the player to make two selections of 6 numbers). In particular it takes on the average 2 Years and 8 months to match four numbers and 25,220 years to win a share in the Jackpot.
- The history of Gosset and Guinness is of course of considerable interest to Irish people, but it is often surprising how few people in an audience know much about it. Perhaps one reason for this is due to the fact that although Gosset or "Student" spent most of his working life in the Guinness breweries in Dublin, he was a member of the Anglo–Irish community in a period of developing Irish independence. Ireland is often known as the land of Saints and Scholars, and so people are always pleasantly surprised to hear about important scientific discoveries in Ireland like that of student's t-test. The story behind Gosset's selection of the pseudonym "student", and the desire of Guinness at the time to keep its scientific discoveries internal to the company, also adds considerable interest to a presentation on statistics (see Boland (2000).

STATISTICS AND THE LAW

Some of the more modern uses of probability and statistics can be seen through examples of probability, statistics and the Law.

- Louis Mountbatten was the 1st Earl Mountbatten of Burma and distinguished himself during WWII as chief of combined British Operations and later as first Sea Lord. He was the last Viceroy of India. During his later years, he often took holidays in County Sligo in Ireland, and he was tragically killed in his yacht by the IRA in Donegal Bay in 1979. Probabilistic evidence was used in Ireland in the murder trial of Mountbatten in the 1980's. Specks of paint removed from the anorak of the accused were collected, and these could not be distinguished from the paint on Mountbatten's boat. Forensic evidence suggested it was extremely unlikely (probability = 1/250,000!) that the accused had not been in contact with Mountbatten's boat. More to the point - it was almost certain (probable beyond a reasonable doubt!) that they had been on the boat in question.
- The Use of Statistical Evidence in Allegations of Exam Cheating usually generates considerable interest (see Boland & Proschan, 1994). Suspicions arose during the final year exams in an Irish university that a breach of security had occurred. In Ireland, the general form of the British education system is used at university level, whereby the final exams are extremely important, and to a large extent determine the quality or class of the degree obtained. There were 17 students in the final year Physics class at this university, and some of the lecturers concerned were worried that a breach of security had taken place after 3 of the 5

final exams had taken place. As a consequence, all of the students were requested to resit versions of Exams 1-3. Although they were not told exactly why this was to be done, the students were given a few days to prepare. Circumstantial evidence and an analysis of the results of these exams led to allegations that two of these students had cheated. A university trial was held, and the statistical analysis provided evidence, which led to their being found guilty. Much of this analysis can be summarized in a simple (x,y) plot (Figure 1). Each point in the plot represents one student in the class. The x co-ordinate represents the average performance on the first attempt at exams 1-3, and the y co-ordinate the second attempt. Points near the diagonal represent students who performed similarly on the first and second set of exams, while points below the diagonal represent those who did poorer on the second set. Peter got average results of (72%, 29%) on the first and second attempts respectively, while Stephen got (64%, 37%). Is this evidence that Peter and Stephen cheated? Interesting discussions can be developed in attempts to assess the value of such evidence.

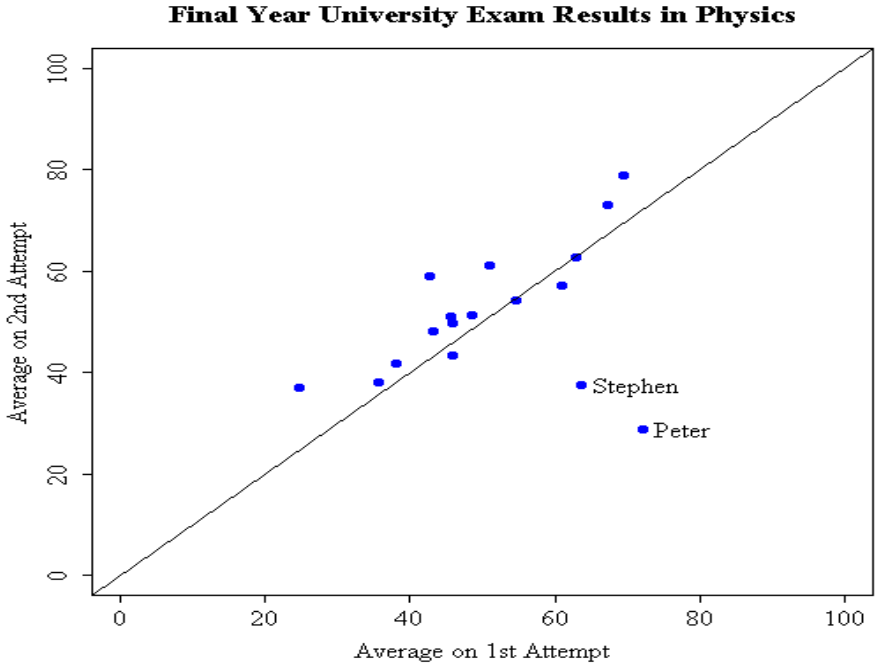

Figure 1. *Relationship between 1st and 2nd Attempt Averages*

THE POWER OF STATISTICAL GRAPICS

The audience response and interest in informative graphical displays can often be surprisingly good, even when using previously unseen techniques like Stem-and-Leaf Diagrams and Box-Plots.

- The ages of Oscar Winning Actresses/Actors from 1928-2000 (see Brown and Davis, (1990)) provides an excellent opportunity to compare two data sets that most people find interesting and for which back to back Stem-and-Leaf plots are very revealing. Somewhat analogously I have gathered data on the age and gender of all presently serving deputies elected in the 1997 Irish General Election in the main house of the Irish Legislature (the Dail). Back to back Stem-and-Leaf plots reveal both differences in the ages and numbers of both genders (in the Irish case females are both considerably fewer and younger). Figure 2 presents an example of graphical display that can be useful in many countries and contexts.

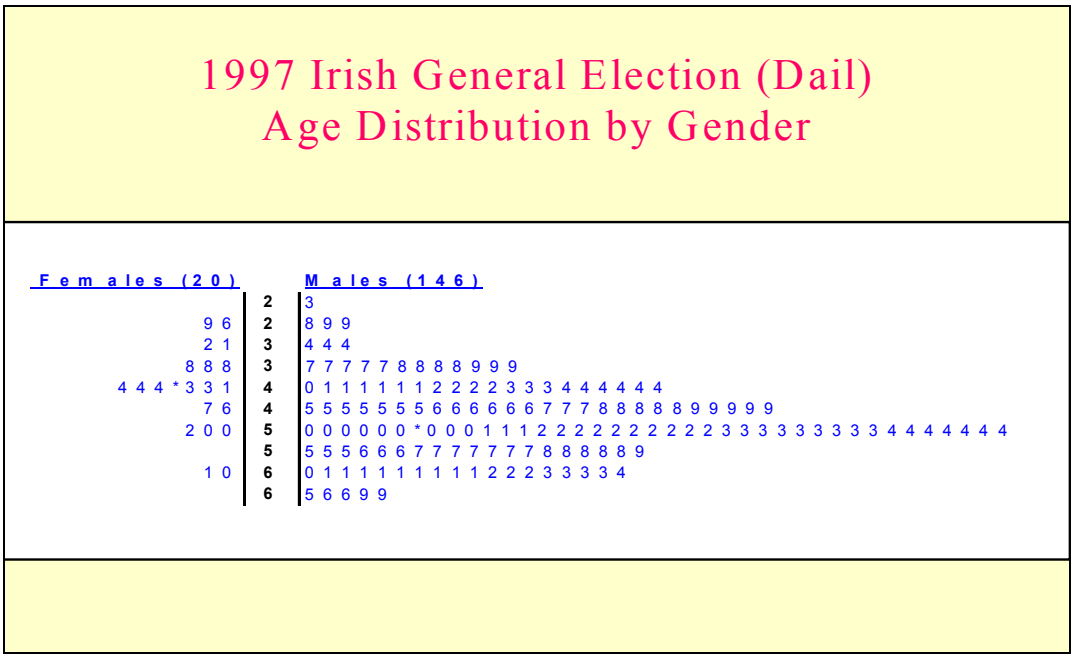

Figure 2. Example of a Stem-and-Leaf Plot

- What day of the week are babies most likely to be born? A box-plot is presented for births in Ireland for each of the seven days of the week in 1994 (Figure 3). This normally generates considerable interest and discussion in spite of the fact that most of the audience have never seen a box-plot before. Although there were on the average 132 births per day in Ireland during the year 1994, the set of seven simultaneous box plots make it clear that there are relatively fewer births during the weekend (particularly Sundays) than in the middle of the week. Interesting discussions can be generated about why this is the case, and also in describing why some days are outliers. Christmas, St Stephen's Day (the day after Christmas which is also known as Boxing Day) and New Year's Day are outliers, which are to some extent to be expected. However I often have fun in teasing out of an Irish audience why Holy Thursday might be an outlier. The fact is that in Ireland the four days after Holy Thursday are for most people holidays, and doctors (obstetricians in particular) like to make good use of them. Easter itself is a major holiday in Ireland. Good Friday is a holiday and the Monday after Easter (known as Easter Monday) is a national holiday commemorating the Easter rising of 1916, which eventually led to Irish independence. I suspect that similar data from other countries will reveal some interesting and special characteristics of a nation's culture.

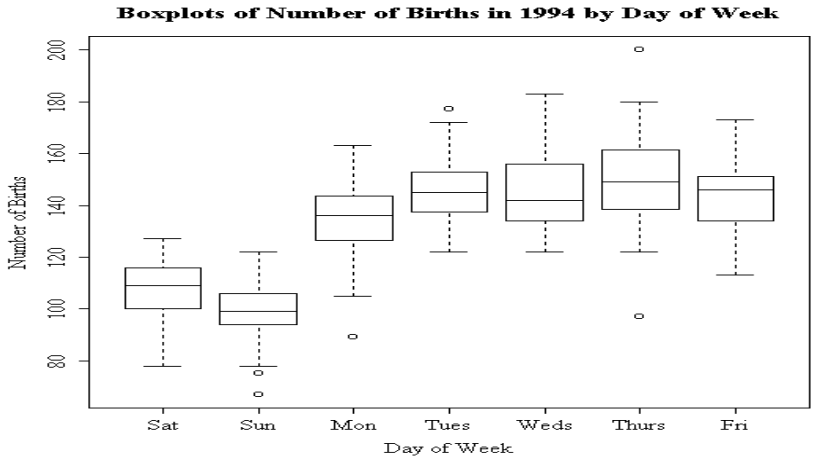

Figure 3. Example of a Box Plot

STATISTICS IN MEDICINE

The story of Florence Nightingale and her "Rose" provides an excellent introduction to the impact of good graphics in epidemiology and community medicine. She very effectively used statistical graphics to show that deaths in the Crimean War were mainly due to sanitary conditions, and not directly from battle deaths as many would have supposed. In a classic graphic, Snow effectively used a map to show that most of the deaths due to cholera clustered around the Broad Street water pump (London, 1854). This had a dramatic effect at the time, but most students today do not even know what cholera is and hence this example does not always make the desired impression on them. The following examples on bovine TB and cancer in Ireland do attract more interest from audiences because of their current and local relevance.

- Bovine tuberculosis has been a serious problem in Ireland for many years, and although cconsiderable effort has been made to eradicate this disease, it still persists. The tuberculin test is the procedure used in Ireland to test an animal for Bovine TB. In this test a vet injects a substance called tuberculin into the animal, and if the skin swells after 72 hours it is declared a reactor (whereby the animal is destroyed and the herd restricted). The main question I put to people in presenting this example is "Do we have a good test for Bovine TB?" Many people (particularly young students) take it for granted that medical tests are very precise and accurate. I use this opportunity to talk about false negatives and false positives and the terminology of specificity and sensitivity of a test. For the Irish tuberculin test, the specificity = P(negative test | healthy animal) = 0.999 and the sensitivity = P(positive test | diseased animal) = 0.90, both of which seem reasonably good. I note also that the incidence of the disease in Ireland is only 0.4%. In spite of these characteristics, I tell my audience that the probability that an animal which tested positive actually has the disease is 0.75 (calculated by using Bayes theorem, although I don't discuss this aspect), and ask them if the average Irish farmer would be happy with this. Given the severe consequences to a farmer of having one of his animals declared a reactor (that is being tested positive), the answer is a definite No! Most farmers in this situation would be thinking of the relatively high chances (25%) of their farm being unnecessarily restricted. I also present a color-coded incidence map of Bovine TB in Ireland by DED (District Electoral Division) based on information gathered in 1999. In order to make geographical comparisons, I stress the need to standardize. In the map displayed, incidence is measured by APT (animals per thousand tests that are positive for tuberculosis). This map of Ireland shows that there is a relatively high incidence of bovine TB near the border with Northern Ireland.
- Cancer is a disease of worldwide concern. Many countries compile considerable data on the incidence of various types of cancer by region, and these can provide good discussion points about statistics and the necessity to standardize rates. For example in Ireland there are regions populated mainly by older people, and hence it is natural to expect higher death rates in these areas. Therefore in order to make fair comparisons between geographical regions, some form of standardization must be used. I do not actually discuss the technical aspects of standardization, but I encourage people to think why this should be done. A geographical map of cancer (excluding non melanoma skin cancer) incidence in Ireland shows that for both males and females, there is a higher incidence of cancer for those living in the main urban areas of Dublin and Cork.

SIMULATION

Simulation is of course an important tool in statistical modelling. Although there is nothing particularly Irish about the package Interactive Probability (see Siegriest, 1996), I have found this piece of software very helpful in providing an introduction to simulation. The module on poker I find to be very instructive in illustrating the concept of long run frequency, and the module on "Let's Make a Deal" always arouses considerable interest. When the audience is relatively small, I have found a discussion about randomness can be greatly enhanced by having them participate in an experiment on estimating the average size of "100 Random Rectangles" (see Burrill & Cobb, 1994).

CONCLUSIONS

This discussion has attempted to emphasize the interest in statistics and statistical thinking that can be generated through the use of examples that are of local interest. These include the use of statistics and the law, statistics on daily births and deaths, and maps of disease incidence that allow fair regional comparisons to be made.

REFERENCES

Boland, P.J. (2000). William Sealy Gosset – alias "student" 1876-1937. In K. Houston (Ed.), *Creators of mathematics: The Irish connection* (pp. 105-112). UCD Press.

Boland, P.J. (1998). Promoting the use of data analysis and statistical projects in Ireland. *Expanding the network: Proceedings of the 5th International Conference on Teaching of Statistics,* Volume 3 (pp. 1127-1134). Singapore.

Boland, P.J., & Nicholson, J. (1996). The statistics and probability curriculum at the secondary school level in the USA, Ireland and the UK. *The Statistician*, *45*, 437-446.

Boland, P.J., & Proschan, M. (1990). The use of statistical evidence in allegations of exam cheating. *Chance*, *3*(3), 10-14.

Boland, P.J., & Pawitan, Y. (1999). Trying to be random in selecting random numbers for lotto. *Journal of Statistical Education*, *7*, 3.

Brown, R., & Davis, G. (1990). A student project: ages of Oscar winning best actors and actresses. *Mathematics Teacher USA, 83*, 2.

Burrill, G., & Cobb, G. (1994). Every one's favourite subject: what's new? *Stats - The Magazine for Students of Statistics*, *Fall,* 18-21.

Siegriest, K. (1996). *Interactive probability*. Wadsworth Publishing Company, Duxbury Press.