# Rescuing Statistics from the Mathematicians.

Mike Bedwell, MA, MSc
Michael_Bedwell@hotmail.com

**Abstract** : Drawing on some 30 years' experience in the UK and Central Europe, the author offers four assertions, three about education generally and the fourth that of the title. There the case is argued that statistics is a branch of logic, and therefore should be taught by experts in such subjects as philosophy and law and not exclusively by mathematicians. Education in both Statistics and these other subjects would profit in consequence.

## Introduction

I am old, I have been working as a jobbing instructor in former Warsaw Pact countries since 1998, and while my main interest has been in statistics, I have in my time taught French, English literature and other subjects in between. With this idiosyncratic introduction, I ask you to indulge my non-academic style, with its shameless use of the first person singular.

I hope to provoke you with four assertions, the most important of which, inspired by Paton (1990) and implied in the title, I shall leave till last. The others are:

## Assertion 1: Keep Technology out of the classroom.

In my UK university, I took for granted that I could turn up to my classroom safe in the knowledge it would be unlocked, warm and well lit, that kind administrators would have informed the students of the where and the when, would have provided sufficient seats, while equally long-suffering technicians would have put in place any equipment I had asked for. But even so, I witnessed too many colleagues going spare when their high-tech aids – over the years ranging from 8mm film to PowerPoint – fail at the critical moment; a particularly embarrassing case was when the material being taught included the reliability of systems comprising components in series.

So in Central Europe, where classroom technicians don't exist and administrators see their roles differently than in the West, I have learned to approach the classroom prepared for the worst; I don't carry candles (but if I did, one of the many smokers among the students would volunteer a light), but I do equip myself with my own chalk and topcoat.  For statistics, I may add a few dice, and a few things the traveller naturally has to hand, such as coins and an opaque bag of variously coloured socks; in their alternative use these are rather more comfortable than the urns and billiard balls beloved of textbook writers. But nothing more.

Lest I be accused of Luddism, I rejoice if outside class hours students can access IT tools. These are invaluable in freeing classroom time for discussion with the students, thereby striving for those higher goals that educationalists wax lyrical about. Indeed, I have gone to the lengths of offering cash to students claiming they cannot afford to access Google, safe in the knowledge that they will be shouted down by their colleagues who know more about available facilities than I do; I have considered a similar approach to embarrass my employers into installing IT tools like a bulletin board.

## Assertion 2: Look West, young teacher.

After my own Oxbridge education, with its frequent essay-writing and infrequent examinations, it came as a shock to teach in Central Europe. For there I have worked in US-sponsored institutions, where the tradition is to require essays only in the liberal arts, but to test frequently, usually with multi-choice questions.

While I still hold reservations about a system that requires, for instance,  a biologist to teach a class without having any say in a putatively prerequisite statistics course, and while I continue to object to the coy practice of calling tests 'quizzes' and essays 'research papers', I have nonetheless become a zealous convert to multi-choice tests. So these I now aim to set every class; I hasten to add that the students are encouraged to work in teams of any size and constitution of their choosing. For me this makes the marking (or 'grading' in American) tractable, even with a class size of over 50. And while it also makes life easy for the class 'passengers', these find themselves shunned over time by the 'workers', and in any case get caught out in the exams, which are sat individually. In Appendix II you will find some sample questions, where I (a) give an example from a non-quantitative subject area, and (b) demonstrate a technique which owes something to Socrates in that it teaches as well as tests. This is essential as after the opening class I aim never to lecture.

Anyone with doubts about American-style testing should tackle one of the Princeton-based Goliaths like the GRE or GMAT; with no other purpose in mind, I submitted myself recently to the latter. I emerged humbled and mentally exhausted, but with increased respect for Princeton's reputation.

**Assertion 3: Exams should be open-book and pre-published.**

No-one would dream of assessing an electrician by asking him/her to write a description of a screwdriver. And most of us are delighted when our doctor consults some authority -- on-line or otherwise – before prescribing our medicine. So I believe the case is overwhelming for open-resource examinations in Statistics and, I suspect, in many other subjects.

Of course, students will use every opportunity to cheat that such examinations seem to present -- and weren't we all students once? This is especially true in Central Europe, where a student whom I would have trusted with my wallet cheerfully boasted to me that academic cheating 'was a national sport'. But my classroom rules in Appendix I and sample questions in Appendix II show ways of meeting this problem; in particular, the theory of Latin squares shows us how to construct an examination comprising a prime number ***n*** questions, each of which has at least ***n*** variants, in such a way that no two out of $\boldsymbol{n}^2$ students will share the same variant in more than one question.

The arguments for pre-publishing exams are less strong, but in my experience the students who have proved leaders in team-testing also do well in their exams, encouraged by my published practice of grading on a relative scale with 100% going to the best. At the same time, the willing donkeys often work out in advance the answers to all variants of all the questions, and also do well, as they surely deserve to.

**Assertion 4: Stand back, Mathematicians.**

Lastly, I advance arguments to support the assertion of my title:

i) As Moore (2002) has extensively illustrated, the most serious problems with statistics in the social sciences, from economics to medicine, have arisen through using an inadequately representative sample, and in particular in coping with non-response. Thus the flawed predictions before the US Presidential elections in 1935 and 1948 have both passed into folklore. If there is a systematic answer to this problem, mathematics is unlikely to provide it.

ii)Mathematics has been described as a branch of logic (Encyclopaedia Britannica, 1998). I hold the same to be true of statistics, so in any taxonomy it should stand side-by-side with maths but not beneath it.

To justify, see what reaction you get the next time you tell someone how you earn your living. Two typical replies are 'there are lies, …', and 'oh yes, I did a bit of that on my ___ course'. In the second case, ask the following 'acid test' question:*Does the p-value derived from statistical testing give the probability of the truth given the evidence, or that of the evidence given the truth?*

The sophisticated response is 'Truth is a difficult philosophical construct; I can reply only if I may substitute "null hypothesis" in its stead'. But, with this modification, how often do you get a straight answer?

In my experience, all too rarely. People who can't answer must have studied from The equivalent of the Biblical book of Proverbs and not of Revelation, yet to see the light they need apply only a little arithmetic, as developing the following table demonstrates:

| What can we learn from the single toss of a coin? | | | |
|---|---|---|---|
| *Null Hypothesis= Prior belief: coin is↓ | *Outcome = evidence↓ | What does the evidence tell us? | Bidirectional *p* value |
| ..conventional – the two sides differ | H (Heads) | Nothing | 100% |
| | T(Tails) | | |
| .. a forgery – both sides the same | H | | 100% |
| | T | | |
| .. a forgery – both sides H | H | | |
| | T | Everything | 0 % |

In this development, we need first to discuss the row and column headings, and the equivalence of the terminology shown *. We could then go on to explore the students' own prior beliefs by asking what they know of Karl Popper's *potential falsification*. And, again before answering, we could explain a *null hypothesis* with the example from criminal law that the accused is innocent until "proved" guilty, but what do we mean by "proof"? In any standard text on statistics, by comparison with one on maths, how often do we encounter this word? None of this discussion needs maths.

But to return to the 'acid test', in deriving the *p* values we are effectively defining 'significance' as the probability of the evidence given our prior belief. We might heartily wish it were the other way round, but once recognized we have the cornerstone of any subsequent discussion of Bayesianism. This topic

Matthews (2005) has not only placed among the 25 'Big Ideas' of current science, but with little maths has explained to the layman.

A further explanation from the table is of an apparent paradox: that the case for abandoning a null hypothesis is stronger when the alternative is unidirectional (the bottom line).

These points can be reinforced by considering two tosses of a coin, as exemplified by test Question 1 in Appendix II. Further, in the context of this question, it is helpful in any more advanced lessons for the teacher to appreciate that, in a sample of two, $t$ is equal to

a) The ratio Sum/Difference, precisely , and
b) tan ($p$% π), where $p$ is the unidirectional $p$ value.

These can be helpful in any later teaching; for example we can readily check the tabulated value of $t = 1.00$ for the 25% unidirectional significance level, and then relate this to the probability of both members of the sample having a median greater than that of the null hypothesis. I would like to continue arguing for thus teaching non-parametric before conventional significance tests, but space precludes. With a final increase to three tosses, we can introduce the concept of *confidence*:

| **What can we learn from an outcomes other than HHH and TTT from tossing a coin three times?** | | | | | |
|---|---|---|---|---|---|
| | Outcome | HHH | Any other | TTT | Sum |
| Statistic→ | Number of Heads | 3 | 1or 2 | 0 | |
| Object is to calculate ↓ | Assumption that the coin is: ↓ | | Probabilities | | |
| Significance | Fair | 1/8 | ¾ | 1/8 | 1 |
| Confidence | Biased as in the sample, i.e. Heads:Tails = 1:2 | 1/27 | 1- (1/27+8/27) =2/3 | 8/27 | 1 |

In developing this table I first introduce the terms *exclusive* and *exhaustive*, so avoiding the need for any further knowledge of probability. I then argue that the two outcomes of either one or two Heads constitute the critical values for a bidirectional significance test at the 1/8 + 1/8 = ¼ probability level, but for a confidence interval of only 2/3 probability, not ¾ .I make this point as many authorities, (eg Upton & Cook, 2001 and Wood, 2003) imply that that sum of the bidirectional *p*-value and confidence interval probability must be one, a generalization explored for more realistic and continuous distributions in Exam Question 3 of Appendix II. Professors Upton and Wood have both been kind enough to debate this issue with me on email, yet have so far left me unrepentant. But I have a hair shirt at the ready!

**Conclusion** We teachers could do better in (1) exploiting new technology, (2) assessing how we assess, and most of all, (3) addressing the question '*Logic', said the Professor....'why don't they teach logic at these schools*?' from C. S. Lewis's *The Lion, the Witch and the Wardrobe.*

**References**


Matthews, Robert. 2005. *25 Big Ideas*. Oneworld publications, 2005.ISBN 10:1-85168-391-7.

Moore, D. 2002 Statistics: *Concepts and Controversies.*

New Encyclopaedia Britannica, 15th edition, 1998. V23, p225 et seq..

Paton, R.1990. Can Statistics be rescued from Mathematics? *Teaching Statistics* 12, 3, 66-69

Upton, G & Cook, I; 2001. *Introducing Statistics*, Oxford UP

Wood, M 2003; *Making Sense of Statistics*, Palgrave Macmillan 1-4039-010704-4


**Appendix I** Instructions to Students, issued in first class.

Welcome. My aim in this course is to minimize the class time we spend in administration and acquiring information, and so maximise the time we spend, often in one-to-one discussion, on the fascinating subject that is Inferential Statistics. While you may not become an expert, you will develop the confidence to hold your own with a professional statistician. And you will learn enough to enlist the help of mathematicians and IT specialists who themselves know no statistics. To these ends, the following will apply:

1. I will make all the documentation you need available on-line on… at least two weeks before you need it. This includes a copy of this Introduction, as well as your weekly tests and Exam. You should print out copies before the class in which you will need them; thus before we next meet, you should print Test 1. You should see me privately if you experience difficulties, financial or otherwise, in achieving this.
2. You should hand in your answers to tests at the end of the appropriate class, and at no other time. You will be encouraged to work on these tests in teams of whatever size or constitution you choose, and to hand in a team answer.

3. You should borrow from any library an introductory statistics text which, on browsing through, you feel comfortable with; this can be and in any language you understand.
4. Your grades-to-date will be published on-line from time to time unless you object that this makes an unwarranted intrusion on your privacy.
5. No late work will be accepted, and there will be no opportunity to sit examinations at other than the prescribed time. This applies no matter how valid your reasons or often you have to be absent from any class, including the first one. However, at the end of term individual oral examinations will be given to students who plead special consideration. At such orals you should present any relevant written medical or other evidence, and be prepared to be interrogated on any of the tests and exams, whether or not you have submitted earlier answers. Other teachers apart from me may be present.
6. 100% in all the tests will earn you a bare pass, but nothing more. But I am happy to report that in the last class the best student got 100 % in the exams, and 99% overall.

**Appendix II** Specimen Test Questions The first time you meet any terms in Italics, check the meaning in a statistics text or on-line equivalent, since in Statistics words often have a more specialized sense than in a general dictionary.

1.Three different cooks A, B and C each measure the diameters of two of their pizzas, and find they are all over-size by the following, in millimetres: A, 0 &12; B, 5 & 15 and C, 4 & 6. All the following are correct, but which two cannot be deduced from this information alone?:

i) A has the biggest *range* ii) B has the biggest *mean*

iii) The range and mean are the most commonly monitored in *Statistical Process Control*

iv) C has the biggest ratio mean/ range, and so has the strongest evidence for taking some corrective action, such as using a smaller pan.

v) The ratio (difference in mean from that of *null hypothesis*): (some measure of *dispersion*) is fundamental to *hypothesis testing*

2.Circle the Null hypotheses in the following:

i) My beliefs are right; show me evidence otherwise.
ii) Your beliefs are wrong; show me evidence otherwise.
iii) In a school there were 10 false fire alarms last term, but the first time the alarm sounds this term the teacher orders “this might be for real –evacuate!”
iv) Elsewhere in the school an examiner orders “this is obviously another false alarm – stay and finish your exam!”
v) In criminal law, the accused is deemed innocent unless strong evidence is found otherwise.

Specimen Questions for an open-book, pre-published Exam

*This exam comprises 7 questions, each with at least 7 variants. On the day of the exam, answer only the variant indicated in pen. Should you accidentally see your neighbour’s paper, you will find he/she shares with you the variant to one question at most.*

1. The following five individuals were prominent during the 20th century: Pope John-Paul II, Joseph Stalin, Adolph Hitler, Margaret Thatcher, and John Lennon. On the day of your exam, you will find one pen-annotated ‘A’, and a second ‘B’. Write an imaginary script for a television interview where A critically interrogates B. [This yields 20 variants]

2. For the sample indicated in the following table, check the calculation of *s* (the best estimate of the population standard deviation when this is otherwise unknown) and calculate the range. The first column gives an example

| Sample | 1 2 3 | 0 1 5 | 1 2 6 | 1 3 8 | 1 2 9 | 0 7 8 | 2 7 9 | 4 8 9 |
|---|---|---|---|---|---|---|---|---|
| *s* | 2 | √28 | √28 | √52 | √76 | √76 | √52 | √28 |
| range | 2 | | | | | | | |

3. Compute the values of the empty cells in the column indicated.

| | Examples* | | | Etc. |
|---|---|---|---|---|
| Sample size n | 40 | 100 | 100 | |
| $\mu$ | 0 | 0 | 0 | |
| $\Sigma$ x | 41.56 | -143.5 | 0 | |
| $\Sigma (x^2)$ | 107.4673 | 1204 | | |
| $\therefore$ $s^2$ | 1.6484 | 10.0816 | | |
| Test statistic z | 5.12 | -4.52 | 1.96 | |
| 95% CI | 200.64 to 201.44 | -2.06 to 0.81 | | |
| Bidirectional *p*- value | <<1% | <<1% | 5% | |

* based on Upton & Cook (*op cit*) p409-410 & 419, after coding. .