STATISTICS FOR THE MATHEMATICALLY CHALLENGED

Mike Bedwell
Wisconsin International University, Kyiv, Ukraine
Michael_Bedwell@hotmail.com

In this paper I argue how we can impart deep understanding of inferential statistics by the early introduction of two approaches that students normally encounter only late in their studies, if at all. These are non-parametric methods and ANOVA. As illustration, we imagine a marathon race where firstly there are just two teams with two runners each, and secondly where there is a third team with two or more runners. We consider various statistics we could use to decide the result, thus discovering the concept of statistical tests in general, and why the chi-square statistic in particular is so important. Following the path of this discovery requires only elementary arithmetic. This is in accordance with the author's belief that students weak in mathematics can learn the fundamentals of inferential statistics at least as readily as a Euclid or a Newton.

INTRODUCTION

The inspiration for this paper came from the need to teach ANOVA. The conventional textbook treatment of this subject seemed to bear more than the usual comparison to a 'cook-book', comprising lots of ugly numbers and formulae, but little of the explanation of the sort to inspire a café cook to develop into a celebrity chef.

So I made it my aim to develop deep understanding of the Kruskal-Wallis technique, the nearest non-parametric equivalent we have to the conventional ANOVA. I recognized the shortcomings of non-parametric methods in general, and in the case of Kruskal-Wallis, of its being limited to one-way analysis. But I also recognized the advantages of non-parametric methods for the beginning student: such methods rest on few of the assumptions which are needed in conventional methods but which are too often confined to the small print. More importantly, rank-sum approaches can lead to a deep understanding of inferential statistics, even by students with little mathematics.

TWO TEAMS OF TWO

We consider two teams X and Y each of two runners, competing in a marathon, where ranks are indeed preferred to times because conditions are less controlled than in, say, a 100-metre dash. The possible outcomes are shown below in the area of Table 1 within the heavy lines; most students will readily recognize that these are 6 in number, since the other 30 cells are unfeasible, the rank of one or both of the runners in X being identical to that of one or both runners in Y. We mark with X+ and Y+ those cells where X and Y achieve their most convincing victories, i.e. where their two runners are placed first and second. If the teams are in fact of equal ability then the chance of each one of these outcomes is 1/6. We note that they can be mutually differentiated by any of the three statistics: x, y, or (x-y), x and y being the total ranks of each team. In the table, we have chosen to show the (x-y) statistic.

If the captain of X will not entertain any thought other than that her team will defeat Y by the greatest possible margin, i.e., (x-y) = +4, then the probability of this happening by a fluke is 1/6. But an unbiased observer may consider that Y is just as likely as X to achieve such a victory, when (x-y) = -4: the *joint probability* of those two outcomes is 2× 1/6=1/3.

Some students may consider the above discussion banal, but fully understood, it illustrates many of the fundamentals of inferential statistics: the meanings of *statistic* (as opposed to *parameter*), of *hypotheses* (*null* & *alternative*), of *exclusive* & *exhaustive,* and of *hypothesis testing*. In particular learners can now resolve the following apparent paradox: that for given evidence, the case for abandoning the null hypothesis in favour of the alternative is stronger in *one-tailed* testing than in *two-tailed.* In other words, we can apparently feel less certain about inferring that an outcome is, on the one hand, <u>different</u> from what would expect under the null hypothesis than, on the other hand, that it is either <u>bigger</u> or <u>smaller</u>. Most fundamentally, students should be able to grasp that what they will learn to call the *p-value*–1/6 and 1/3 above–expresses the probability of the evidence given the null hypothesis, much as we would wish it were the other way

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

round. The failure of so many conventional texts to emphasize this crucial principle surely amounts to criminal negligence.

Table 1.

Ranks of the two members of each team	X	1,2	1,3	1,4	2,3	2,4	3,4	….	5,6
Y	Sum of those ranks x→ y↓	3	4	5	5	6	7		11
1,2	3						4 Y+		+8 Y+
1,3	4					2			
1,4	5				0				
2,3	5			0					
2,4	6		-2;						
3,4	7	-4: X+							4 Z+
….									
5,6	11	-8 X+					-4 Z+		

THREE TEAMS OF TWO

To investigate what happens when the race is entered by a third team Z, also of two, we have to expand the above table. The students should be allowed to try this on their own, perhaps with the hint that the number of both rows and columns will increase from 6 to 15. To help, I have in my more compassionate moods issued them with a sheet of A3 paper, and urged them, as I always do, to work with pencil+eraser rather than pen. For reasons that will become clear later, only 7 of those 15 rows and columns have been shown in the Table 1.

Students should begin by putting a cross in all the unfeasible cells; this will include all 30 of the blank cells in the corner of the table used in the two-team case. While this process might seem tedious, students usually find it enjoyable–perhaps it has the same appeal as Sudoku, which engages people who would deny any interest in statistics? Further, students will without encouragement collude, especially if told to expect a total of 90 feasible cells. These should now be filled in with the (x-y) statistic; many will come to recognize the symmetry of the table; indeed, it is possible to implement this exercise by examining only those 45 cells to one side or other of the 'northwest-southeast' diagonal.

We can now consider how suitable (x-y) remains as the statistic. The two cells already marked ± 4 continue to correctly indicate the greatest possible triumphs of X and Y over each other, and two of the additional such cells indicate where Z scores its total triumphs Z+. But what about the other cells where (x-y) =± 4, and what can we infer from the two cells marked ±8, the new maximum and minimum values? Well, other of the ± 4 cells denote no total victory, while we observe that ±8 can only occur when one or other of X and Y scores 7 and the other 11, so Z must score 21-7-11 =3, indicating a total victory denoted by Z+. For we come to appreciate that for a team to describe its victory as 'total', it must not only score 3, but one of the other teams must score 11, when it thus suffers a 'total defeat'. Such appreciation is valuable to the learning process, and may be reinforced by showing x and y to be equally unsatisfactory as statistics. For application rather than learning, however, we may conclude we need a statistic that is some function of z as well as of x and y.

This is the cue to introduce the chi-square test. Again, once shown a sample calculation, students can be left to get on with calculating the values of this statistic for all the feasible cells, and again this will not prove as onerous as it might at first seem; the fact that the variables are integers means that even a hand calculator is unnecessary. If the sample calculation is for one of the 6 'total victories' already identified, then the hint can be given that these 6 cells, and these only, have a chi-square of 4.6. From tabulated values of chi-square, this corresponds to a 2-sided *p*-value of 10% (2df), rather more than the true value of 6/90, or about 7%. But if we return to two-team case and repeat the exercise, we find the approximation still worse; this suggests, correctly, that the tabulated values of chi-squared should be relied on only for larger values of N.

The expected results are summarized in Table 2. There is no need to confront the students with the formulae given from combination theory; these are included for the teachers to mull over, and to enable them if they wish to explore what happens when n_z is increased to 3 or more. It will then be found that the results can be checked against the 5% critical values published for the Kruskal-Wallis test. But again there is no need for students to know this by name, any more than they need to recognize the *sign* or *Mann-Whitney* tests as such, even though the teacher should appreciate that these were in effect being used in case with N=4

Table 2.

Total Number of Runners	N	4	6
Number of runners in each team	n_x	2	2
	n_y	2	2
	n_z	-	2
R = Number of different columns and rows, i.e. of different rank sums for each of x and y	N!/ {2(N-2)!}	6	15
Number of cells	R(R-1)	30	210
F= Number of feasible cells	N!/{4(N-4)!}	6	90
Statistic: Difference of ranks (x-y)	Minimum	-4	-8
	Maximum	+4	+8
B=Number of Cells with these maxima & minima		2	2
p- value, 2-tail test, from this Statistic	B/F	33%	2%
Statistic: Chi-sq	Max calculated value	1.60	4.6
C=Number of cells with this Max		2	6
p- value, 2-tail test, from this Statistic	C/F	33.3%	6.7%
p- value, 2-tail test, published value.	From Chi-sq table	20% (1df)	10% (2df)

CONCLUSION

Although the Kruskal-Wallis test is rarely covered in conventional introductory text books, the process of understanding it leads to the unraveling of most of the fundamental principles of inferential statistics. Yet this vital learning needs none of the mathematics with which such texts are commonly burdened.

REFERENCES

Bedwell, M. (2009). Rescuing Statistics from the Mathematicians. *Proceedings of Conference on Modelling in Mathematics Education, Dresden, September 2009* (pp. 52-57).