Parabola Volume 13, Issue 3 (1977)

SIMPSON'S PARADOX, OR THE BOWLER'S COMPETITION

Alan Fekete*

A recent letter to the magazine "New Scientist" drew attention to the following unexpected fact: if two bowlers perform equally well in one innings, but differently in the next, their overall performances may still be equal! In cricket a bowler's performance is determined by the number of runs scored against him, and the number of wickets taken, and it is the ratio of these numbers that matters. Suppose in one innings, bowlers Smith and Jones capture every wicket, each taking 5 for 60, but in the next innings Smith gets 1 for 18, while Jones takes 3 for 57. Clearly Smith is the better bowler with a ratio of 1:18, compared with Jones' 1:19. However the total figures reveal Smith with 6 for 78 and Jones with 9 for 117, each a ratio of 1:13.

This is as far as the letter writer took the matter, but the situation can be even more confusing. Take another two bowlers, Black and White. In the first innings, Black takes 3 for 72, while White takes 3 for 75, and in the second innings they take 1 for 16 and 3 for 51 respectively. Thus in each innings Black shows himself to be the better bowler. When the whole match is considered, however, White with 6 for 126 clearly outscores Black with 4 for 88!

In this form, the problem will only worry cricket record keepers, but this is just an example of a problem called Simpson's Paradox, which was first discussed in the 1930's. At the Mathematics Day for 6th Form students this year, Dr Hoetzl of the N.S.W. Institute of Technology discussed the paradox and its implications. In this article I will use his example and explanation, together with my solution, if it can be called that, of the problem.

Consider an examination for 400 students, which is held in two large halls, each seating 200 people. A statistician, no doubt a male chauvinist pig, wanted to prove that boys were better than girls at this subject. He found that in Hall "A", 7 out of the 98 girls failed (or 7.1%) while only 6 out of 102 boys (6.6%) failed. Encouraged, he found that in Hall "B" 6 out of 45 girls failed (13.3%) and 18 out of 155 boys failed (11.6%). This, the statistician thought, clearly shows that a smaller percentage of boys failed than girls. However, when we examine the total

Alan is in Year 12 at Sydney Grammar School.

figures we find that 13 girls out of 143 failed (9.09%) but 24 out of 257 boys failed (9.33%). We can show these figures like this:

	Hall A		Hall B		Total	
	Failed	Sat	Failed	Sat	Failed	Sat
Men	6	102	18	155	24	257
Women	7	98	6	45	13	143

If this seems unimportant, suppose that instead of men and women we had patients treated with a drug and patients not treated, that passing and failing represented being cured or dying and that the two halls were two hospitals. Each hospital would report that the drug helped cure the disease, but considering the complete data we find that it clearly causes more deaths than it saves lives.

In the case of the students, Dr Hoetzl showed a possible explanation: A jackhammer was going outside Hall B, where more of the men were sitting for the exam. This pulled the men's performance down below the women's. However, how can we be sure that each hall wasn't made of two parts, one well lit, the other in the dark, with women over-represented in the poorly lit part? Perhaps, in each section of each hall the women were better, but in each hall they were hurt by being in the dark. Any figures we see could be composed of two parts, each of which gives a different picture from the total.

Having hopefully made you doubt the value of all statistical claims about drugs, or H.S.C. marks, or cigarette smoking, I would like to tell you how I have managed to live with the problem. It seems to me that this paradox merely makes clear that one should never forget that statistics can never show any causal connection between things. They never show that White is a better bowler than Black, or that women are better than men in a subject, or that cigarette smoking causes lung cancer, but only that cigarette smoking is associated with lung cancer, that women are linked to doing well, or that White is linked to taking wickets. The link may be something like "People who worry, smoke, and are also likely to get sick", "women tended to be in the quiet hall, where people did better" or "White did more of his bowling in the second innings, when the wicket was better". This doesn't prevent us acting on data, but it does mean we must be on the lookout for connections of this kind instead of assume a direct, causal link.

Questions to investigate:

1. With the cricket scores, consider that we "added" ratios of '5 to 60' and '1 to 18' and got '6 to 78'. Why don't we add fractions like this:

$$\frac{5}{60} + \frac{1}{18} = \frac{5+1}{60+18}$$

[For 6th Form Students: Why, then, do we add complex numbers and vectors like this:

$$(5 + 60i) + (1 + 18i) = 6 + 78i \quad \text{and} \quad (5,60) + (1,18) = (6,78)?$$

Note that both complex numbers and fractions are introduced in the syllabus as ordered pairs.]

2. Show that this method of adding fractions is equivalent to finding a weighted mean. What are the weightings?

3. What can we do with ratios? See a Mathematics Dictionary under Proportion (if a:b = c:d, then a+b:b = c+d:d, a+b:a–b = c+d:c–d and so on).

4. Show that if men and women are equally distributed between two halls, no paradox can occur.

Appendix – how to find examples of a paradox: first split men and women unevenly between the rooms (1), then choose different failure rates for the rooms (2), calculate the totals (3), then reduce the failure rate among whichever group failed more, but keeping the total higher (4). For example:

		Room 1		Room 2		Total	
		Fail	Sat	Fail	Sat	Fail	Sat
(1)	Men		50		150		200
	Women		150		50		200
(2)	Men	5	50	30	150	35	200
(3)	Women	15	150	10	50	25	200
(4)	Men	1	50	29	150	30	200
	Women	15	150	10	50	25	200