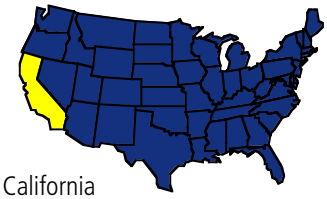
SIMPSON'S PARADOX

1

BERKELEY SEX BIAS CASE

David Freedman, Robert Pisani and Roger Purves. "Statistics" (3rd edition). W.W. Norton, 1998, p. 19.



State of California

One of the best known real life examples of Simpson's paradox occurred when the University of Berkeley was sued for bias against women who had applied for admission to graduate schools there. The admission figures for the fall of 1973 showed that men applying were more likely than women to be admitted, and the difference was so large that it was unlikely to be due to chance.

	applicants	admitted
men	2590	46 %
women	1835	30 %

However when examining the individual departments, it was found that no department was significantly biased against women. In fact, most departments had a "small but statistically significant bias in favour of women".

department	men		women	
	applicants	admitted	applicants	admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

Show that Simpson's paradox occurs and explain why.

The groups of applicants for admission are **differently composed**.

In the ones to departments A and B the women form a small minority (12 % and 4 % respectively). Exactly here their percentage of admission is higher than the one of the men.
Also in the group of applicants to department D and F, where men and women balance each other, the chances are slightly in favour of the women.

Finally in the group of applicants to department C and E, where the women form a **majority** (65 % and 67 % respectively), their percentage of admission is slightly lower than the one of the men. **But exactly these two departments provide a number of failures big enough to shift the women's average below the one of the men.**

2

BATTING AVERAGES

Ken Ross. "A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans" Pi Press, 2004

A common example of Simpson's Paradox involves the batting averages of players in professional baseball. It is possible for one player to have a higher batting average than another player during a given year, and to do so again during the next year, but he has a lower batting average when the two years are combined. This phenomenon can occur when there are large differences in the number of at-bats between the years. (The same situation applies to calculating batting averages for the first half of the baseball season, and during the second half, and then combining all of the data for the season's batting average.)
A real-life example is provided by Ken Ross and involves the batting average of two baseball players, Derek Jeter and David Justice, during the baseball years 1995 and 1996:

	1995	**1996**	**combined**
Derek Jeter	12/48	183/582	195/630
David Justice	104/411	45/140	149/551

Show that Simpson's paradox occurs and explain why.

First the relative frequencies of the batting averages have to be calculated:

	1995	**1996**	**combined**
Derek Jeter	0.250	0.314	0.310
David Justice	0.253	0.321	0.270

Derek Jeter's batting average in 1995 and 1996 is lower than David Justice's, but his average over both of the years is higher! This is **Simpson's paradox**.

It is caused by the **different number of at-bats each year**. In 1995 Justice plays 9 times as many bats as Jeter but both of them have a low average and therefore a low number of bats which do not contribute much to the total.
In 1996 Jeter plays 4 times as many bats as Justice and both of the players have a high average. So the influence of the many bats by Jeter on the total is big!