

A particularly intriguing type of confounding occurs when a confounding variable, once exposed, completely turns around the observed association. This phenomenon goes by the name of Simpson's paradox, although it is not a true paradox.

A very high profile example (made *Science* magazine), occurred in the early seventies in a study of graduate school admissions to Berkeley.

In fall quarter, 1973, there were 8,442 men who applied for admission to graduate school, and 4,321 women. About 44% of the men and 35% of the women were admitted. Since admissions are made separately for each major, the admissions data were broken down to find out which majors were discriminating against women. Major by major, there did not seem to be any bias against women. The data for the six largest majors, accounting for over one third of the total number of applicants, is given below (from [FPP]):

Major	Male applicants		Female applicants	
	Total	% admitted	Total	% 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	373	6	341	7
Total	2691	44	1835	30

Over 50% of the men applied to the first two majors, which were relatively easy to get into. Over 90% of the women applied to the last four majors, which were much harder to get into. The confounding variable is the major applied to; it is hidden in a comparison of the overall admission rates.

Here is another example, from [Rad], on the relationship between race and the imposition of the death penalty for convicted first-degree murderers.

Race of defendant	Death penalty		
	Yes	No	% Yes
White	19	141	11.88
Black	17	149	10.24

The imposition of the death penalty is about the same rate for white and black defendants, in fact, slightly higher for white defendants. However, when the race of the victim is taken into account, a quite different picture emerges:

Race of defendant	White victim			Black victim		
	Death penalty			Death penalty		
	Yes	No	% Yes	Yes	No	% Yes
White	19	132	12.58	0	9	0
Black	11	52	17.46	6	97	5.83

Here is a third example from [Wag]. In the mid-seventies, the tax rate in the United States was lowered for every category of income. However, people's income increased, meaning that they started moving to the higher rates.

Table 2. Total Income and Total Tax (in thousands of dollars), and Tax Rate for Taxable Income Tax Returns, by Income Category and Year

Adjusted Gross Income	1974			1978		
	Income	Tax	Tax Rate	Income	Tax	Tax Rate
under \$ 5,000	41,651,643	2,244,467	.054	19,879,622	689,318	.035
\$ 5,000 to \$ 9,999	146,400,740	13,646,348	.093	122,853,315	8,819,461	.072
\$ 10,000 to \$14,999	192,688,922	21,449,597	.111	171,858,024	17,155,758	.100
\$ 15,000 to \$99,999	470,010,790	75,038,230	.160	865,037,814	137,860,951	.159
\$ 100,000 or more	29,427,152	11,311,672	.384	62,806,159	24,051,698	.383
Total	880,179,247	123,690,314		1,242,434,934	188,577,186	
Overall Tax Rate			.141			.152

Source: [Wag]

References

- [FPP] Freeman, D., Pisani, R., and Purves, R. (1998). *Statistics* (3rd edition). WW Norton.
- [Rad] Radelet, M. (1981). Racial characteristics and imposition of the death penalty, *American Sociological Review*, **46**, 918–927. There is a related article in *Chance News* 4.04 from the *New York Times*, February 24, 1995.
- [Wag] Wagner, C.H. (1982). Simpson's paradox in real life, *The American Statistician*, **36**, 46–48.