The Yule-Simpson paradox

In statistics, *Effect Reversal*, sometimes known as the Yule-Simpson's paradox, describes a phenomenon in which statistical characteristics existing in subsets of a group of data changes when these subsets are combined. However, if there is some structure in the multivariate data set (certain marginal or conditional independences), we can avoid the paradox.

Example 1.

Murder sentences in Florid (see the The New York Times paper).

Table 1: Florida 1973-77 Death Penalty and Victim/Offender Race Comparison

Victim/Offender Race	Number of Offenders	Death Sentences	Probability of Death Penalty
B kills W	286	48	.168
W kills W	2146	72	.034
B kills B	2320	11	.005
W kills B	111	0	.000

Notations:

Let M be the binary variable indicating the race of the murderer (B/W).

Let S be the binary variable indicating if death sentence is given (y=capital punishment, n=other punishment).

Let V be the binary variable indicating the race of the victim (B/W).

In this case, we consider V as the latent variable (denoted by W in the sequel), M as the predictor variable (denoted by X in the sequel), and S as the response variable (denoted by Y in the sequel).

Observation:

The probability of a black offender being given a death penalty: $P(S = \mathrm{Y}, M = \mathrm{B}) = 0.0226$, and the probability of a white offender being given a death penalty: $P(S = \mathrm{Y}, M = \mathrm{W}) = 0.0319$

This result shows that, in general, black offenders are less likely to be sentenced to death.

However, when given the race of the victim, we have different results, as illustrated by the table above.

We can observe that the court sanctions more seriously killing white victim than a black one. Since the majority of the victims of black murderers is black, and that of white murderers is white, a larger number of white murderers is sentenced to death in summary.

Observe the joint distribution of $M - S$ in the slices according to the values of V is different than the marginal joint distribution of $M - S$ (when we sum for the two values of V).

Example 2.

Danish women

This example is taken from an investigation of 237 Danish women performed by the Gallup Institute. Relation between the use of physical punishment, political affiliation and childhood experience of physical punishment in 237 Danish women is shown in the table below.

Table 2: Results obtained from the investigation.

Childhood experience and use of physical punishment(PP)		Political Affiliation		
		Left	Soc.Dem.	Right
Has experienced PP	Uses PP	12	27	58
	Does not use PP	7	28	30
Has not experienced PP	Uses PP	9	5	9
	Does not use PP	19	15	18

$Y = U$ (usage of physical punishment y/n), $X = E$ (childhood experience of physical punishment y/n), $W = A$ (party affiliation, L/S/R).

Observation:

Table 3: Tables describing marginal associations between childhood experience (E) of physical punishment, use (U) of physical punishment and political affiliation (A)

	U	
E	yes	no
yes	97	65
no	23	52

	A		
E	l	s	r
yes	19	55	88
no	28	20	27

	A		
U	l	s	r
yes	21	32	67
no	26	43	48

When we made a U versus A cross-tabulation, and estimated the marginals, the 2×3 table manufactured as the product of the marginal probabilities, was very close to the original layers of the $2 \times 2 \times 3$ table, both in the $E = y$ and $E = n$ cases. We can also make a χ^2 test on the two 2×3 tables.

Table 4: Results from the χ^2 tests for independence

Null Hypothesis (H_0)	$U \perp\!\!\!\perp A$	$U \perp\!\!\!\perp A \vert E = \text{y}$	$U \perp\!\!\!\perp A \vert E = \text{n}$
Degree of Freedom	2	2	2
χ^2 Value	5.25	4.08	0.42
Significance	0.072	0.130	0.810
Conclusion (7.5%)	Reject	Do not reject	Do not reject

Conclude for the marginal and conditional independence of A and U (conditioned on E)!