

# Outbreak of Simpson's Paradox

by  
Paul Mullenix, Ph.D.  
SystatS Consulting Sdn. Bhd.

No, Simpson's Paradox does not refer to how O.J. could be acquitted. Rather, Simpson's Paradox is a famous example of the sometimes non-intuitive nature of statistics.

The outbreak of this paradox occurred in case of percentages met for quality problem actions in response to complaints. The symptoms of the disorder appear in the following set of data.

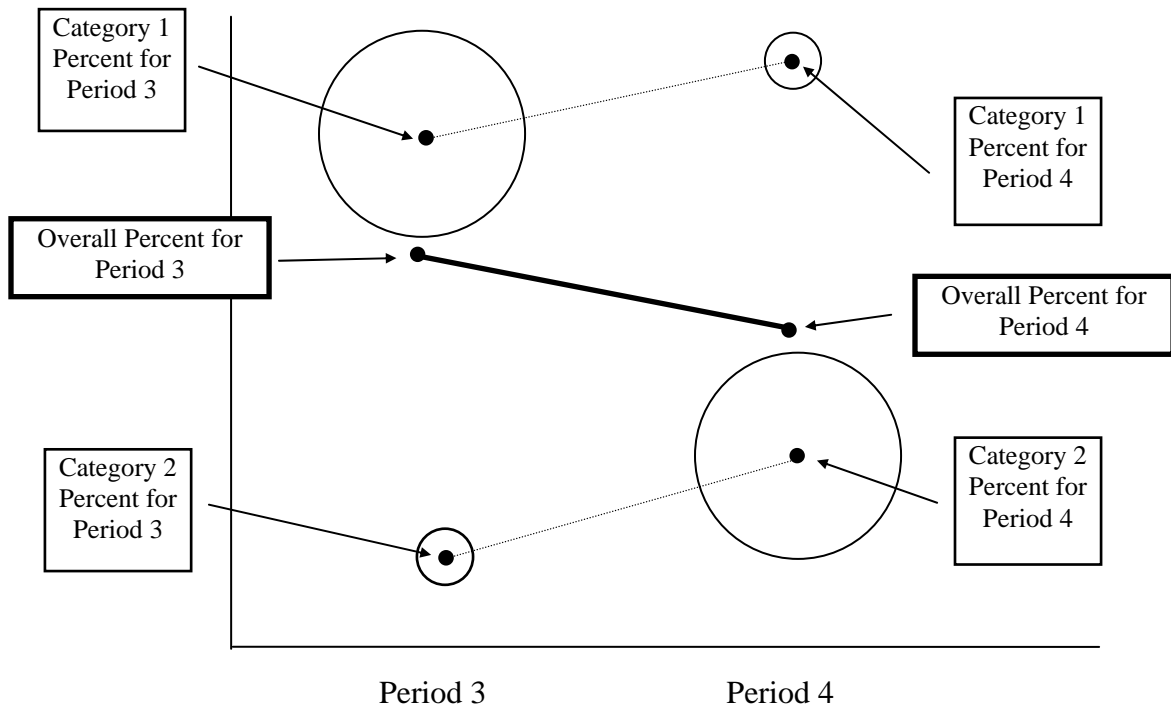
Action Category		Period 3	Period 4
Electrical	Met	40	25
	Cases	41	25
	Percent	98%	100%
Failure Analysis	Met	25	17
	Cases	27	17
	Percent	93%	100%
Corrective Action	Met	17	18
	Cases	20	19
	Percent	85%	95%
Visual	Met	1	10
	Cases	9	24
	Percent	11%	42%
Total	Met	83	70
	Cases	97	85
	Percent	86%	82%

Notice that each individual category percentage increased from period 3 to period 4, yet the overall total percentage went down! How is it possible that each category improved in period 4, but the period 4 overall total percentage was worse than period 3?

The diagnosis of this ailment is known as Simpson's Paradox. Although it is named after E. H. Simpson who discussed the paradox in 1951 in the *Journal of the Royal Statistical Society* Vol. B13: pp. 238-241, it should more appropriately be named after G. U. Yule, who discussed it as early as 1903 in *Biometrika* Vol. 2: pp. 121-134.

The difficulty in the above example is that the low percentage in the Visual category during period 3 is weighted very little (9 out of the 97 total cases) in the overall percentage for period 3. However, the low percentage for the Visual category during period 4 is weighted heavily (24 out of the 85 total cases). The paradoxical quality comes from our minds trying to weight each category effect equally to calculate the overall effect. If each category percentage increases, equal weighting suggests that the overall total percentage should increase also.

The following diagram illustrates this weighting difficulty for a simpler case of only two categories. Notice that the circular areas represent the numbers of cases or weights for each category. Each individual category improved as shown by the dashed lines, yet the overall percentage went down as shown by the solid line. This overall decline is due to the heavy weighting of the high value for Category 1 in period 3 and the heavy weighting of the low value for Category 2 in period 4.



The cure to eliminate this apparent paradox is to be aware of the different bases used in the calculation of the percentages. The individual category percentages use the total number of cases for that category, while the overall period percentages use the total number of cases for that period from all categories combined.

With every percentage, two pieces of information need to be considered. One is the percentage itself which expresses the relationship of the number of some occurrence to the total number. The second is the magnitude of the total number. A percentage by itself may be misleading.

*This is the reason that pie charts may be misleading, they only focus on the percentages and not on the total number of cases. When is the last time you saw a monthly pie chart drawn with the size of pie reflecting the total number of cases?*

As a treatment for the case at hand, we may report the delinquencies as a percent of the period total, not as a percent of the category total. The following table illustrates the idea.

Action Category		Period 3	Period 4
Electrical	Number of Delinquent	1	0
	Percent of Total Cases	1%	0%
Failure Analysis	Delinquent	2	0
	Percent of Total Cases	2%	0%
Corrective Analysis	Delinquent	3	1
	Percent of Total Cases	3%	1%
Visual	Delinquent	8	14
	Percent of Total Cases	8%	17%
Total	Delinquent	14	15
	Cases	97	85
	Percent	14%	18%

This clearly shows that all categories except Visual improved in period 4, but the real good news is that the overall number of action commitments continues to decline from 97 to 85! The rise in the overall percentage can be seen as due to the Visual category where 17% of the action commitments were delinquent for period 4; up from 8% delinquent cases for period 3. Notice also that the percentages for the categories in a given period now add up to the overall delinquent percentage for that period.

Misleading percentages are contagious. Similar phenomena occur elsewhere in manufacturing industries when the mix of products is changed. We sometimes explain a change in overall yield as due to a difference in mix. Thus, in many applications the weighting factors turn out to be very important when looking at the overall picture. For those interested budding statisticians who find the technical details infectious, the explanation for Simpson's Paradox lies in a violation of the weak consistency condition as reported by Haunsperger and Saari in *American Statistician* Vol. 45, No. 3, pp. 252-255.

So the next time you are exposed to Simpson's Paradox you need not worry, you have just received your inoculation by being able to recognize and understand the effect of varying weighting factors and different percentage bases.