Simpson's paradox

From Wikipedia, the free encyclopedia

Simpson's paradox, or the **Yule–Simpson effect**, is a paradox in probability and statistics, in which a trend appears in different groups of data but disappears or reverses when these groups are combined. It is sometimes given the descriptive title **reversal paradox** or **amalgamation paradox**.[1]

This result is often encountered in social-science and medical-science statistics[2][3] and is particularly confounding when frequency data is unduly given causal interpretations.[4] The paradoxical elements disappear when causal relations are brought into consideration.[5] Many statisticians believe that the mainstream public should be informed of the counter-intuitive results in statistics such as Simpson's paradox.[6][7] Martin Gardner wrote a popular account of Simpson's paradox in his March 1976 Mathematical Games column in *Scientific American*.[8]

Edward H. Simpson first described this phenomenon in a technical paper in 1951,[9] but the statisticians Karl Pearson et al., in 1899,[10] and Udny Yule, in 1903,[11] had mentioned similar effects earlier. The name *Simpson's paradox* was introduced by Colin R. Blyth in 1972.[12]
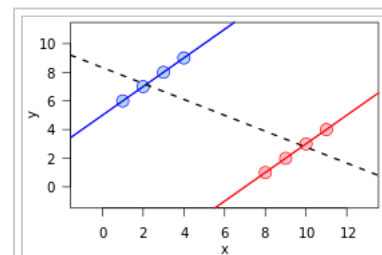
Simpson's paradox for quantitative data: a positive trend (——, ——) appears for two separate groups, whereas a negative trend (- - -) appears when the groups are combined.

Contents

Examples

UC Berkeley gender bias

One of the best-known examples of Simpson's paradox is a study of gender bias among graduate school admissions to University of California, Berkeley. 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.[13][14]

	Applicants	Admitted
Men	8442	**44%**
Women	4321	35%

But when examining the individual departments, it appeared that six out of 85 departments were significantly biased against men, whereas only four were significantly biased against women. In fact, the pooled and corrected data showed a "small but statistically significant bias in favor of women."[14] The data from the six largest departments is listed below.

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	373	6%	341	**7%**

The research paper by Bickel et al.[14] concluded that women tended to apply to competitive departments with low rates of admission even among qualified applicants (such as in the English Department), whereas men tended to apply to less-competitive departments with high rates of admission among the qualified applicants (such as in engineering and chemistry). The conditions under which the admissions' frequency data from specific departments constitute a proper defense against charges of discrimination are formulated in the book *Causality* by Judea Pearl.[4]

Kidney stone treatment

This is a real-life example from a medical study[15] comparing the success rates of two treatments for kidney stones.[16]

The table below shows the success rates and numbers of treatments for treatments involving both small and large kidney stones, where Treatment A includes all open surgical procedures and Treatment B is percutaneous nephrolithotomy (which involves only a small puncture). The numbers in parentheses indicate the number of success cases over the total size of the group. (For example, 93% equals 81 divided by 87.)

	Treatment A	Treatment B
Small stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

The paradoxical conclusion is that treatment A is more effective when used on small stones, and also when used on large stones, yet treatment B is more effective when considering both sizes at the same time. In this example, the "lurking" variable (or **confounding variable**) is the severity of the case (represented by the doctors' treatment decision trend of favoring B for less severe cases), which was not previously known to be important until its effects were included.

Which treatment is considered better is determined by an inequality between two ratios (successes/total). The reversal of the inequality between the ratios, which creates Simpson's paradox, happens because two effects occur together:

1. The sizes of the groups, which are combined when the lurking variable is ignored, are very different. Doctors tend to give the severe cases (large stones) the better treatment (A), and the milder cases (small stones) the inferior treatment (B). Therefore, the totals are dominated by groups 3 and 2, and not by the two much smaller groups 1 and 4.
2. The lurking variable has a large effect on the ratios; i.e., the success rate is more strongly influenced by the severity of the case than by the choice of treatment. Therefore, the group of patients with large stones using treatment A (group 3) does worse than the group with small stones (groups 1 and 2), even if the latter used the inferior treatment B (group 2).

Based on these effects, the paradoxical result is seen to arise by suppression of the causal effect of the severity of the case on successful treatment. The paradoxical result can be rephrased more accurately as follows: When the less effective treatment (B) is applied more frequently to less severe cases, it can appear to be a more effective treatment.

Low birth weight paradox

The low birth weight paradox is an apparently paradoxical observation relating to the birth weights and mortality of children born to tobacco smoking mothers. As a usual practice, babies weighing less than a certain amount (which varies between different countries) have been classified as having low birth weight. In a given population, babies with low birth weights have had a significantly higher infant mortality rate than others. Normal birth weight infants of smokers have about the same mortality rate as normal birth weight infants of non-smokers, and low birth weight infants of smokers have a much lower mortality rate than low birth weight infants of non-smokers, but infants of smokers overall have a much higher mortality rate than infants of non-smokers. This is because many more infants of smokers are low birth weight, and low birth weight babies have a much higher mortality rate than normal birth weight babies.[17] However, some other causes of low birth weight carry a much higher infant mortality rate.

Batting averages

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 each year for a number of years, but to have a lower batting average across all of those years. 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[18] and involves the batting average of two baseball players, Derek Jeter and David Justice, during the years 1995 and 1996:[19]

	1995		1996		Combined	
Derek Jeter	12/48	.250	183/582	.314	195/630	**.310**
David Justice	104/411	**.253**	45/140	**.321**	149/551	.270

In both 1995 and 1996, Justice had a higher batting average (in bold type) than Jeter did. However, when the two baseball seasons are combined, Jeter shows a higher batting average than Justice. According to Ross, this phenomenon would be observed about once per year among the possible pairs of interesting baseball players. In this particular case, the Simpson's Paradox can still be observed if the year 1997 is also taken into account:

	1995		1996		1997		Combined	
Derek Jeter	12/48	.250	183/582	.314	190/654	.291	385/1284	**.300**
David Justice	104/411	**.253**	45/140	**.321**	163/495	**.329**	312/1046	.298

The Jeter and Justice example of Simpson's paradox was referred to in the "Conspiracy Theory" episode of the television series *Numb3rs*, though a chart shown omitted some of the data, and listed the 1996 averages as 1995.

Correlation between variables

Simpson's paradox can also arise in correlations, in which two variables appear to have (say) a positive correlation towards one another, when in fact they have a negative correlation, the reversal having been brought about by a "lurking" confounder. Berman et al.[20] give an example from economics, where a dataset suggests overall demand is positively correlated with price (that is, higher prices lead to *more* demand), in contradiction of expectation. Analysis reveals time to be the confounding variable: plotting both price and demand against time reveals the expected negative correlation over various periods, which then reverses to become positive if the influence of time is ignored by simply plotting demand against price.

Description

Suppose two people, Lisa and Bart, each edit articles for two weeks. In the first week, Lisa fails to improve the only article she edited, and Bart improves 1 of the 4 articles he edited. In the second week, Lisa improves 3 of 4 articles she edited, while Bart improves the only article he edited.

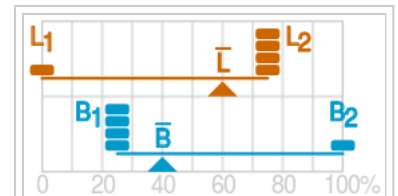|	Week 1	Week 2	Total
Lisa	0/1	3/4	**3/5**
Bart	**1/4**	**1/1**	2/5

Illustration of Simpson's Paradox; the upper figure represents Lisa's contribution and the lower one Bart's. The left bars represent the first week, the right bars the second week; the triangles indicate the combined percentage of good contributions (weighted average). While each of Bart's bars show greater success than Lisa's, Lisa's combined rate is higher because she improved a greater ratio relative to the quantity edited.

Both times Bart improved a higher percentage of articles than Lisa, but the actual number of articles each edited (the bottom number of their ratios, also known as the *sample size*) were not the same for both of them either week. When the totals for the two weeks are added together, Bart and Lisa's work can be judged from an equal sample size; i.e., the total number of articles edited by each. Looked at in this more accurate manner, Lisa's ratio is higher and, therefore, so is her percentage. Also when the two tests are combined using a weighted average, overall, Lisa has improved a much higher percentage than Bart because the quality modifier had a significantly higher percentage. Therefore, like other paradoxes, it only appears to be a paradox because of incorrect assumptions, incomplete or misguided information, or a lack of understanding a particular concept.

	Week 1 quantity	Week 2 quantity	Total quantity *and* weighted quality
Lisa	0%	75%	**60%**
Bart	**25%**	**100%**	40%

This imagined paradox is caused when the percentage is provided but not the ratio. In this example, if only the 25% in the first week for Bart was provided but not the ratio (1:4), it would distort the information and so cause the imagined paradox. Even though Bart's percentage is higher for the first and second week, when two weeks of articles is combined, overall Lisa had improved a greater proportion, 60% of the 5 total articles. Lisa's proportional total of articles improved exceeds Bart's total.

Vector interpretation

Simpson's paradox can also be illustrated using the 2-dimensional vector space.[21] A success rate of $\frac{p}{q}$ (i.e., *successes/attempts*) can be represented by a vector $\vec{A} = (q, p)$, with a slope of $\frac{p}{q}$. A larger slope, meaning a steeper vector direction, represents then a more successful week. If two rates $\frac{p_1}{q_1}$ and $\frac{p_2}{q_2}$ are combined, as in the examples given above, the result can be represented by the sum of the vectors (q_1, p_1) and (q_2, p_2), which according to the parallelogram rule is the vector $(q_1 + q_2, p_1 + p_2)$, with slope $\frac{p_1 + p_2}{q_1 + q_2}$.

Vector interpretation of Simpson's paradox for the Lisa and Bart example
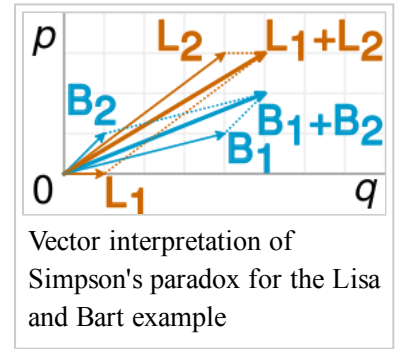
Simpson's paradox says that even if a vector $\vec{L_1}$ (in light brown in the figure) has a smaller slope than another vector $\vec{B_1}$ (in blue), and $\vec{L_2}$ has a smaller slope than $\vec{B_2}$, the sum of the two vectors $\vec{L_1} + \vec{L_2}$ can still have a larger slope than the sum of the two vectors $\vec{B_1} + \vec{B_2}$, as shown in the example.

Implications for decision making

The practical significance of Simpson's paradox surfaces in decision making situations where it poses the following dilemma: Which data should we consult in choosing an action, the aggregated or the partitioned? In the Kidney Stone example above, it is clear that if one is diagnosed with "Small Stones" or "Large Stones" the data for the respective subpopulation should be consulted and Treatment A would be preferred to Treatment B. But what if a patient is not diagnosed, and the size of the stone is not known; would it be appropriate to consult the aggregated data and administer Treatment B? This would stand contrary to common sense; a treatment that is preferred both under one condition and under its negation should also be preferred when the condition is unknown.

On the other hand, if the partitioned data is to be preferred a priori, what prevents one from partitioning the data into arbitrary sub-categories (say based on eye color or post-treatment pain) artificially constructed to yield wrong choices of treatments? Pearl[4] shows that, indeed, in many cases it is the aggregated, not the partitioned data that gives the correct choice of action. Worse yet, given the same table, one should sometimes follow the partitioned and sometimes the aggregated data, depending on the story behind the data, with each story dictating its own choice. Pearl[4] considers this to be the real paradox behind Simpson's reversal.

As to why and how a story, not data, should dictate choices, the answer is that it is the story which encodes the causal relationships among the variables. Once we explicate these relationships and represent them formally, we can test which partition gives the correct treatment preference. For example, if we represent causal relationships in a graph called "causal diagram" (see Bayesian networks), we can test whether nodes that represent the proposed partition intercept spurious paths in the diagram. This test, called "back-door," reduces Simpson's paradox to an exercise in graph theory (see page 7 of [22])

Psychology

Psychological interest in Simpson's paradox seeks to explain why people deem sign reversal to be impossible at first, offended by the idea that an action preferred both under one condition and under its negation should be rejected when the condition is unknown. The question is where people get this strong intuition from, and how it is encoded in the mind. Simpson's paradox demonstrates that this intuition cannot be derived from either classical logic or probability calculus alone, and thus led philosophers to speculate that it is supported by an innate causal logic that guides people in reasoning about actions and their consequences. Savage's sure-thing principle[12] is an example of what such logic may entail. A qualified version of Savage's sure thing principle

can indeed be derived from Pearl's *do*-calculus[4] and reads: "An action *A* that increases the probability of an event *B* in each subpopulation C_i of *C* must also increase the probability of *B* in the population as a whole, provided that the action does not change the distribution of the subpopulations." This suggests that knowledge about actions and consequences is stored in a form resembling Causal Bayesian Networks.

Probability

A paper by Pavlides and Perlman presents a proof, due to Hadjicostas, that in a random **2 × 2 × 2** table with uniform distribution, Simpson's paradox will occur with a probability of exactly $^{1}/_{60}$[23] A study by Kock suggests that the probability that Simpson's paradox would occur at random in path models (i.e., models generated by path analysis (statistics)) with two predictors and one criterion variable is approximately 12.8 percent; slightly higher than 1 occurrence per 8 path models.[24]

Related concepts

- Ecological fallacy (and ecological correlation)
- Modifiable areal unit problem
- Prosecutor's fallacy
- Anscombe's quartet

References

1. I. J. Good, Y. Mittal (June 1987). "The Amalgamation and Geometry of Two-by-Two Contingency Tables". *The Annals of Statistics*. **15** (2): 694–711. doi:10.1214/aos/1176350369. ISSN 0090-5364. JSTOR 2241334.
2. Clifford H. Wagner (February 1982). "Simpson's Paradox in Real Life". *The American Statistician*. **36** (1): 46–48. doi:10.2307/2684093. JSTOR 2684093.
3. Holt, G. B. (2016). Potential Simpson's paradox in multicenter study of intraperitoneal chemotherapy for ovarian cancer. (http://jco.ascopubs.org/content/34/9/1016.1.full) Journal of Clinical Oncology, 34(9), 1016-1016.
4. Judea Pearl. *Causality: Models, Reasoning, and Inference*, Cambridge University Press (2000, 2nd edition 2009). ISBN 0-521-77362-8.
5. Kock, N., & Gaskins, L. (2016). Simpson's paradox, moderation and the emergence of quadratic relationships in path models: An information systems illustration. (http://cits.tamiu.edu/kock/pubs/journals/2016JournalIJANS_ModJCve NetCorrp/Kock_Gaskins_2016_IJANS_SimpPdox.pdf) International Journal of Applied Nonlinear Science, 2(3), 200-234.
6. Robert L. Wardrop (February 1995). "Simpson's Paradox and the Hot Hand in Basketball". *The American Statistician*, **49 (1)**: pp. 24–28.
7. Alan Agresti (2002). "Categorical Data Analysis" (Second edition). John Wiley and Sons ISBN 0-471-36093-7
8. Gardener, Martin (March 1979). "MATHEMATICAL GAMES: On the fabric of inductive logic, and some probability paradoxes" (PDF). *Scientific American*. **234**: 119. Retrieved 28 February 2017.
9. Simpson, Edward H. (1951). "The Interpretation of Interaction in Contingency Tables". *Journal of the Royal Statistical Society, Series B*. **13**: 238–241.
10. Pearson, Karl; Lee, Alice; Bramley-Moore, Lesley (1899). "Genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses". *Philosophical Transactions of the Royal Society A*. **192**: 257–330. doi:10.1098/rsta.1899.0006.
11. G. U. Yule (1903). "Notes on the Theory of Association of Attributes in Statistics". *Biometrika*. **2** (2): 121–134. doi:10.1093/biomet/2.2.121.
12. Colin R. Blyth (June 1972). "On Simpson's Paradox and the Sure-Thing Principle". *Journal of the American Statistical Association*. **67** (338): 364–366. doi:10.2307/2284382. JSTOR 2284382.
13. David Freedman, Robert Pisani, and Roger Purves (2007), *Statistics* (4th edition), W. W. Norton. ISBN 0-393-92972-8.
14. P.J. Bickel, E.A. Hammel and J.W. O'Connell (1975). "Sex Bias in Graduate Admissions: Data From Berkeley" (PDF). *Science*. **187** (4175): 398–404. doi:10.1126/science.187.4175.398. PMID 17835295.
15. C. R. Charig; D. R. Webb; S. R. Payne; J. E. Wickham (29 March 1986). "Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy". *Br Med J (Clin Res Ed)*. **292** (6524): 879–882. doi:10.1136/bmj.292.6524.879. PMC 1339981. PMID 3083922.

16. Steven A. Julious; Mark A. Mullee (3 December 1994). "Confounding and Simpson's paradox". *BMJ*. **309** (6967): 1480–1481. doi:10.1136/bmj.309.6967.1480. PMC 2541623. PMID 7804052.

17. Wilcox Allen (2006). "The Perils of Birth Weight — A Lesson from Directed Acyclic Graphs". *American Journal of Epidemiology*. **164** (11): 1121–1123. doi:10.1093/aje/kwj276. PMID 16931545.

18. Ken Ross. "*A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans (Paperback)*" Pi Press, 2004. ISBN 0-13-147990-3. 12–13

19. Statistics available from Baseball-Reference.com: Data for Derek Jeter (http://www.baseball-reference.com/j/jeterde0 1.shtml); Data for David Justice (http://www.baseball-reference.com/j/justida01.shtml).

20. Berman, S. DalleMule, L. Greene, M., Lucker, J. (2012), "Simpson's Paradox: A Cautionary Tale in Advanced Analytics (http://www.statslife.org.uk/the-statistics-dictionary/2012-simpson-s-paradox-a-cautionary-tale-in-advanced -analytics)", *Significance*.

21. Kocik Jerzy (2001). "Proofs without Words: Simpson's Paradox" (PDF). *Mathematics Magazine*. **74** (5): 399. doi:10.2307/2691038.

22. Pearl, Judea (December 2013). "Understanding Simpson's paradox" (PDF). *UCLA Cognitive Systems Laboratory, Technical Report R-414*.

23. Marios G. Pavlides & Michael D. Perlman (August 2009). "How Likely is Simpson's Paradox?". *The American Statistician*. **63** (3): 226–233. doi:10.1198/tast.2009.09007.

24. Kock, N. (2015). How likely is Simpson's paradox in path models? (http://cits.tamiu.edu/kock/pubs/journals/2015Jou rnalIJeC/Kock_2015_IJeC_SimpPdox.pdf) International Journal of e-Collaboration, 11(1), 1–7.

Bibliography

- Leila Schneps and Coralie Colmez, *Math on trial. How numbers get used and abused in the courtroom*, Basic Books, 2013. ISBN 978-0-465-03292-1. (Sixth chapter: "Math error number 6: Simpson's paradox. The Berkeley sex bias case: discrimination detection").

External links

- How statistics can be misleading - Mark Liddell (http://ed.ted.co m/lessons/how-statistics-can-be-misleading-mark-liddell)—TED-Ed video and lesson.

> Wikimedia Commons has media related to *Simpson's paradox*.

- Stanford Encyclopedia of Philosophy: "Simpson's Paradox (http://plato.stanford.edu/entries/paradox-sim pson/)" – by Gary Malinas.
- Earliest known uses of some of the words of mathematics: S (http://jeff560.tripod.com/s.html)
  - For a brief history of the origins of the paradox see the entries "Simpson's Paradox" and "Spurious Correlation"
- Pearl, Judea, ""The Art and Science of Cause and Effect. (http://bayes.cs.ucla.edu/LECTURE/lecture_sec 1.htm)" A slide show and tutorial lecture.
- Pearl, Judea, "Simpson's Paradox: An Anatomy" (http://bayes.cs.ucla.edu/R264.pdf) (PDF)
- Simpson's Paradox Visualized (http://vudlab.com/simpsons/) - an interactive demonstration of Simpson's paradox.
- Pearl, Judea, "The Sure-Thing Principle" (http://ftp.cs.ucla.edu/pub/stat_ser/r466.pdf) (PDF)
- Short articles by Alexander Bogomolny at cut-the-knot:
  - "Mediant Fractions. (http://www.cut-the-knot.org/blue/Mediant.shtml)"
  - "Simpson's Paradox. (http://www.cut-the-knot.org/Curriculum/Algebra/SimpsonParadox.shtml)"
- The Wall Street Journal column "The Numbers Guy" (https://online.wsj.com/article/SB12597074455307 1829.html) for December 2, 2009 dealt with recent instances of Simpson's paradox in the news. Notably a Simpson's paradox in the comparison of unemployment rates of the 2009 recession with the 1983 recession, by Cari Tuna (substituting for regular columnist Carl Bialik).
- How to resolve Simpson's paradox? (http://stats.stackexchange.com/questions/78255/how-to-resolve-sim psons-paradox) question on statistics Q&A site CrossValidated

Retrieved from "https://en.wikipedia.org/w/index.php?title=Simpson%27s_paradox&oldid=771647702"

Categories: Probability theory paradoxes │ Statistical paradoxes │ Causal inference │ 1951 introductions

---

This page was last modified on 22 March 2017, at 19:13.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

- Contact Wikipedia
- Developers
- Cookie statement