

Statistisches Edutainment: Die Umkehrung von Vergleichen¹

LAWRENCE M. LESSER, UNIVERSITY OF TEXAS AT EL PASO UND DENNIS K. PEARL,
PENNSYLVANIA STATE UNIVERSITY

¹ Original: Lesser, L. M. & Pearl, D. K. (2019).
Statistical edutainment: Reversing comparisons.
Teaching Statistics, 41, 118–122.
Übersetzung und Bearbeitung:
BIRGIT GRIESE, PADERBORN

Zusammenfassung: *Amüsante und zugleich lehrreiche Elemente sind einfach zu finden und können zur Steigerung der Motivation bei Studierenden oder Schüler*innen eingesetzt werden. Wir stellen Ressourcen und Vorgehensweisen für einen speziellen Inhalt vor: das Simpson-Paradoxon.*

1 Einleitung

„Edutainment“ ist eine Wortschöpfung aus „education“ (Bildung) und „entertainment“ (Unterhaltung) und ist zu einem Bestandteil der deutschen Sprache geworden. Damit werden in diesem Artikel amüsante und zugleich lehrreiche Elemente bezeichnet, wie beispielsweise Gedichte, Songs, Cartoons, Rätsel und Witze, die sich auf Lerninhalte aus der Statistik beziehen, im Unterricht erprobt wurden (inklusive Überlegungen zum bestmöglichen Einsatz), und für deren Einsatz im Unterricht weder viel Zeit noch komödiantisches Talent erforderlich sind. Unsere Anmerkungen speisen sich aus Jahren der Unterrichtserfahrung, empirischer Forschung und unserer Tätigkeit als Kuratoren für eine internationale Sammlung an Spaß-Elementen zu den Grundlagen der Statistik unter <https://www.CAUSEweb.org/fun>. Diese oft besuchte Sammlung beinhaltet aktuell über 800 Elemente aus 11 Kategorien: Kunst, Cartoons, Spiele, Witze, Zauberei, Gedichte, Rätsel, Zitate, Songs, Geschichten und Videos. Die Sammlung kann durchsucht werden, es gibt Filter für Kategorien, inhaltliche Themen und Stichworte. Derartige Elemente können verbreitete Bedenken (wie in Lesser et al., 2013, geschildert) auf Seiten der Lehrkräfte zerstreuen oder sogar in das Gegenteil umkehren.

2 Das Simpson-Paradoxon

A propos Umkehrungen, wir können dieses Prinzip auf ein bestimmtes statistisches Phänomen anwenden, das Simpson-Paradoxon. Typischerweise wird es als die Tatsache beschrieben, dass die Richtung eines Vergleichs von zwei Variablen sich umkehrt, wenn man verschiedene Untergruppen vereint. Alternativ kann auch die Richtung eines Vergleichs von zwei Variablen teilweise oder komplett umgekehrt werden, wenn diese in Untergruppen aufgeteilt werden, also eine dritte Variable hinzugenommen wird. In einem hypothetischen Beispiel werden zwei entzündungshemmende Medikamente für schnelle Schmerzlinderung an 200 Patient*innen (die dem Test zugestimmt haben) für Medikament A und 400 Patient*innen für Medikament B getestet. Dabei wird Wirksamkeit (Schmerzlinderung nach 30 Minuten) bei 8 % der Patient*innen mit Medikament A und 8,5 % der Patient*innen mit Medikament B beobachtet. Wenn die Patient*innen jedoch nach Altersgruppen (unter 50 Jahren, mindestens 50 Jahre alt) aufgeteilt werden, könnte sich herausstellen, dass sich dieser Vergleich in einer oder beiden Gruppen umkehrt, wie zum Beispiel in Tabelle 1 dargestellt.

Das Simpson-Paradoxon (das im GAISE College Report von 2016, S. 10, 38, erwähnt wird) ist für Schüler*innen der Sekundarstufe mit ihrem Wissen über Anteile gut zugänglich, aber dennoch nuanciert und komplex genug, um auch Studierende zu interessieren, vgl. Lesser (2001) und Schield (2006) für Hintergründe und vielfältige Repräsentationen des Simpson-Paradoxon als Tabelle, Kreisdiagramm, Wägebücke, Trapez, BK-Plot, Einheitsquadrat, mit Determinanten und in der Vektorgeometrie. Dieses Thema ist auch ein gutes Beispiel dafür, dass sich amüsante und zugleich lehrreiche Elemente zu praktisch jedem Thema der Statistik finden lassen.

Um nicht nur zu erkennen, *dass* ein Phänomen auftreten kann, sondern auch, *unter welchen Umständen* dies

	Alter < 50 Jahre, Stichprobenumfang	Wirksamkeit, Anzahl und Anteil, < 50 Jahre	Alter ≥ 50 Jahre, Stichprobenumfang	Wirksamkeit, Anzahl und Anteil, ≥ 50 Jahre	Wirksamkeit, gesamt
Medikament A	150	9 (6 %)	50	7 (14 %)	16 von 200 = 8 %
Medikament B	100	4 (4 %)	300	30 (10 %)	34 von 400 = 8,5 %

Tab. 1: Hypothetische Daten für die Wirksamkeit von zwei Medikamenten, aufgeteilt nach Altersgruppen

passiert, ist aktives Explorieren erforderlich, beispielsweise indem man eine Lerneinheit wie die zum Werfen eines Balls aus zerknülltem Papier (Gou & Zhang, 2017) einsetzt, oder eine App wie <http://www.math.usu.edu/~schneit/Statlets/Simpson's%20Paradox/> (Schneider, 2013) oder https://rstudio.aws.science.psu.edu:3838/Boast/Multivariable_Topics/Paradox/.

3 Das Simpson-Paradoxon als Song

Wie in Tabelle 1 dargestellt, ist die greifbarste Art, das Simpson-Paradoxon darzustellen, eine Tabelle mit Zahlen. Ein Song hingegen kann das Phänomen vorstellen und dazu motivieren, eine Datentabelle zu erstellen oder zu untersuchen.

Die Autoren beauftragten den Singer/Songwriter Monty Harper, der für seine lehrreichen wissenschaftlichen Songs preisgekrönt wurde, für ihr vom NSF ermöglichtes Projekt SMILES einen Song zum Simpson-Paradoxon zu schreiben. Dort ist ein konkretes Beispiel im Songtext aufgeführt, siehe <https://www.causeweb.org/cause/resources/fun/songs/no-one-counted-simpsons-paradox>.

Die ersten beiden Strophen beschreiben die Situation eines gemischten Kinderchors, bei dem die durchschnittliche Größe der Jungen und die durchschnittliche Größe der Mädchen im Vergleich zum Vorjahr anstiegen, aber dennoch die durchschnittliche Gesamtgröße abnimmt. Die letzte Strophe erklärt, wie das möglich ist (hier in deutscher Übersetzung):

Wir hatten mehr Mädchen dieses Jahr,
Und weniger Jungs dieses Jahr,
Und Mädchen sind meistens kleiner
Als Jungs, normalerweise.

Also sind unsere Mädchen gewachsen,
Und unsere Jungs sind gewachsen,
Aber der Chor ist geschrumpft,
Und wir haben was gelernt.

Um ihr Verständnis zu zeigen, können die Schüler*innen eigene Daten passend zum Song konstruieren. Tabelle 2 zeigt eine solche Tabelle für einen Chor aus 10 Kindern pro Jahr.

	Anzahl Mädchen (durchschn. Größe)	Anzahl Jungen (durchschn. Größe)	Gesamt
aktuell	8 (155 cm)	2 (168 cm)	10 (157,6 cm)
Vorjahr	2 (152 cm)	8 (165 cm)	10 (162,4 cm)

Tab. 2: Größe von Jungen und Mädchen im Chor

Für weitere Motivation und Beschäftigung können die Schüler*innen sich mit dem Song in einer interakti-

ven Version (im Netz unter https://www.CAUSEweb.org/smiles/songs/simpsons_paradox, erstellt im NSF-Projekt SMILES) beschäftigen, indem sie Teile des Textes nach Aufforderung ergänzen, die dann im Playback besonders hervorgehobene Teile des Songs werden, ähnlich wie in dem Spiel Mad Lips, mit dem englische *Phrasal Verbs* geübt werden können (und das in Trumppower, 2010, für den Statistikerunterricht umgeschrieben wurde). So können Schüler*innen ihr Verständnis zeigen, indem sie einen neuen Kontext konstruieren (d. h. ihr Verständnis auf ein anderes Szenario als Körpergrößen und Mädchen/Jungen in einem Chor anwenden).

Ein weiteres Beispiel in der CAUSEweb.org-Sammlung (zu finden unter <https://www.CAUSEweb.org/cause/resources/fun/videos/simpsons-paradox>) ist ein einminütiges Video einer Parodie des Musikvideos zum Hit „Another one bites the dust“ von Queen, von der US-amerikanischen Lehrerin Mary McLellan von der Aledo High School in Texas. Der Song enthält die Pointe des Simpson-Paradoxons und im Video wird eine Tabelle mit passenden Zahlen gezeigt.

Obwohl (noch) nicht alle Schüler*innen Statistik lieben, mögen alle Musik, so dass der Einsatz von Songs, um die Beschäftigung mit statistischen Phänomenen zu motivieren, überzeugt. Wenn Lehrkräfte aus Unsicherheit zögern, diesen Ansatz zu verfolgen, können sie vielleicht durch Forschungsergebnisse überzeugt werden (z. B. <http://singaboutscience.org/wp/educating/research/>). Lehrkräfte, die sich nicht in der Lage sehen, selbst zu singen, können Songs aus CAUSEweb herunterladen und im Unterricht abspielen.

4 Das Simpson-Paradoxon im Cartoon

Ein anderer populärer Ansatz für lehrreichen Spaß sind Cartoons, wie zahlreiche Studien (z. B. Özdoğru & McMorris, 2013), Cartoon-Bücher (z. B. Klein & Dabney, 2013) und Unterrichtsentwürfe zu Cartoons (z. B. Lesser, 2018) belegen. Abbildung 1 zeigt eine Zeichnung des Londoner Cartoonisten John Landers, basierend auf einer Idee von Dennis Pearl. Nachdem sie sich kurz an dem Cartoon erfreut haben, können Schüler*innen gefragt werden: „Wenn die meisten Männer im Publikum den Witz mochten, und die meisten Frauen im Publikum ebenfalls, ist es dann überhaupt möglich, dass das Publikum insgesamt den Witz nicht gut fand?“ Um die richtige Antwort „nein“ zu dieser Frage zu finden, müssen die Schüler*innen aktiv einen Erkenntnisprozess durchlaufen, um zu erkennen, dass (a) der Wert eines gewichteten Mittels zwischen den Mittelwerten der beiden Untergruppen

liegen muss, und (b) das Simpson-Paradoxon eine dritte Variable benötigt, aber in dem Cartoon nur zwei vorkommen („Geschlecht“ und „Spaß an dem Witz“). Und da wir gerade beim Unmöglichen sind: Ein weiterführendes anspruchsvolles Problem ist die Frage, warum das, was Friedlander und Wagon (1993) ein „doppeltes Simpson-Paradoxon“ nennen, nicht auftreten kann.



„Ich verstehe nicht, warum das Publikum meinen Witz zum Simpson-Paradoxon hasst. Ich habe ihn separat an den Männern und Frauen im Publikum ausprobiert, und die mochten ihn jeweils.“

Abb. 1: Cartoon zum Simpson-Paradoxon, <https://www.CAUSEweb.org/cause/resources/fun/cartoons/Simpsons-Comic>, farbig unter wileyonlinelibrary.com

Ein weiterer Cartoon von Landers zum Simpson-Paradoxon (Abb. 2) wurde zur CAUSEweb-Sammlung hinzugefügt, nachdem die Bildunterschrift des Erstautors zum Gewinner des CAUSEweb-Wettbewerbs gekürt worden war. (Dieser monatliche Wettbewerb <https://www.CAUSEweb.org/cause/caption-contest/> steht seit drei Jahren allen Schüler*innen, Studierenden und Lehrkräften offen.) Die visuelle Darstellung eines Maisfeldes kann Lernenden helfen, die sogenannten „Cornfield-Bedingungen“ zu memorieren, weniger bekannte notwendige, nicht hinreichende Bedingungen für das Auftreten des Simpson-Paradoxons. Man kann sie als „minimale Effektgröße eines potentiellen Störfaktors, die notwendig ist, um einen beobachteten Zusammenhang unter der Annahme, dass dieser völlig unberechtigt ist, zu erklären“ beschreiben (vgl. Schield, 1999, S. 2). Im Beispiel des Wirksamkeitstests von zwei Medikamenten A und B aus Tabelle 1 beträgt die Gesamtdifferenz zwischen A und B nur 0,5 % (A = 8,5 %, B = 8 %), allerdings ist

das Simpson-Paradoxon möglich, denn es gibt einen Störfaktor („Alter“), der eine größere Differenz von ca. 5,4 % aufweist ($\text{Alter}_{<50} = 5,2\%$, $\text{Alter}_{\geq 50} = 10,57\%$). Im Fall von Tabelle 2, bei der wir uns für Größenunterschiede von Chormitgliedern in aufeinanderfolgenden Jahren interessieren, ist das Simpson-Paradoxon möglich, weil der Unterschied von 4,8 cm zwischen den Jahren (162,4 cm und 157,6 cm) kleiner ist als der Unterschied von 11,2 cm zwischen den Geschlechtern (165,5 cm für Jungen, 154,4 cm für Mädchen).



Cornfield's conditions ruled out Simpson's paradox from a potential corn-founder.

Abb. 2: Gewinner des CAUSEweb-Wettbewerbs vom Juni 2019 für die beste Bildunterschrift, <https://www.CAUSEweb.org/cause/resources/fun/cartoons/corn-fields-condition>, farbig unter wileyonlinelibrary.com. Das englische Original enthält das Wortspiel zu Cornfield (Namensgeber der Cornfield-Bedingungen)/wörtliche Bedeutung „Maisfeld“ und bezieht sich auf die Lautähnlichkeit von *corn-founder* (*corn* = Mais, *founder* = Gründer/Errichter) mit *confounder* (Störfaktor), zu Deutsch etwa „Laut den Cornfield-Bedingungen kann das Simpson-Paradox als potentieller Störfaktor ausgeschlossen werden.“

5 Das Simpson-Paradoxon in Witzen

Da Cartoons implizit stets einen verbalen und/oder einen visuellen Witz enthalten, liegt es nahe, auch Witze auf ihr Potential für lehrreichen Spaß zu betrachten. Tatsächlich kann man gut und gerne behaupten, das Simpson-Paradoxon selbst sei aufgebaut wie ein Witz. Wie http://michaelnielsen.org/reinventing_explanation/ ausführt, „beginnt das Simpson-Paradoxon mit Voraussetzungen, die uns dazu verleiten, ein bestimmtes Ergebnis zu erwarten, und dann, bääm!, kommt eine plötzliche Wende, und wir verstehen, dass man die Situation auch völlig anders auffassen kann. Es ist eine Herausforderung, das richtig darzustellen, weil das Publikum die Pointe vielleicht nicht sofort kapiert, außer sie sind vorbereitet worden. Es erfordert eine sorgsam geplante Vorgehensweise“ (aus dem englischen Original übertragen). Die Verwendung von Witzen und Humor in der Statistik ist

allgemein (über die Grenzen der CAUSEweb-Sammlung hinaus) beforscht worden, und eine sehr gute englischsprachige Website ist unter <http://my.ilstu.edu/~gcramsey/Gallery.html> zu finden.

Auch in einer speziellen Variante des Simpson-Paradoxons, des sogenannten Will-Rogers-Phänomens, ist Humor (und auch Wahrheit) verborgen. Hier bewirkt die Verschiebung eines Elements von einer Teilmenge in eine andere, dass sich das arithmetische Mittel in beiden Teilmengen erhöht, während es in der Gesamtmenge unverändert bleibt. Eine populäre Anwendung machte Rogers in der 1930ern, als er scherzte: „Als die Okies (Einwohner von Oklahoma) Oklahoma verließen und nach Kalifornien zogen, haben sie das durchschnittliche Intelligenzniveau in beiden Staaten erhöht“, ähnlich wie Neuseelands damaliger Premierminister Robert Muldoon ein halbes Jahrhundert später auf Kosten Australiens (Orsman & Moore, 1988). Lernende können selbst aktiv mit Hilfe von Zahlen herausfinden, wie das passieren kann, indem man sie auffordert, zwei Mengen von Zahlen $\{a, b\}$ und $\{c, d, e\}$ zu finden, wobei c kleiner als d und e aber größer als a und b ist. Sie können dann einfach nachvollziehen, dass die Verschiebung von c in die andere Menge das arithmetische Mittel beider Teilmengen erhöht. Weiteres Ausprobieren zeigt, dass c nicht notwendigerweise die kleinste Zahl der zweiten Menge ist. Ein interessanter Realitätsbezug kann im Bereich der Medizin gefunden werden: Verbesserte Krebsnachweismetho-

den ermöglichen frühere Diagnosen und führen damit zu einer erhöhten Klassifikation der Befunde in ernsteren Krankheitsstadien. Diese Verschiebung der Befunde zwischen den Krebsstadien verbesserte die Überlebensrate in allen Stadien (Sormani, 2009).

6 Das Simpson-Paradoxon in Gedichten

Es ist sicherlich nicht überraschend, dass ein weitere Variante von Spaß mit Statistik in Gedichten zu finden ist, denn Gedichte waren bereits auf den Titelseiten von Fachzeitschriften (z. B. <http://www.radstats.org.uk/journal/issue117/>), wurden in Blogs für Mathematik-Gedichte vorgestellt (z. B. <https://poetrywithmathematics.blogspot.com/2014/10/abc-of-statistics.html>) und tauchten in Statistik-Zeitschriften auf (Champkin, 2011). Das folgende Gedicht (Lesser, 2010) zum Simpson-Paradoxon wird hier mit freundlicher Genehmigung des *Mathematical Intelligencer* abgedruckt und ist um eine deutsche Übersetzung ergänzt.

Nach dem Lesen des Gedichts, kann sich die Klasse an den folgenden Fragen versuchen:

- Zeige, dass $3/8 > 1/3$, $2/3 > 3/5$, aber $5/11 < 4/8$ ist, und weise damit nach, dass der Vergleich zwischen dem Autor und seiner Rivalin umgekehrt wird, wenn man ihre Einreichungen bei den verschiedenen Zeitschriften zusammenfasst.

„Confounded“ (L. Lesser)

3 of 8 poems
I submitted
to the classic journal
were accepted,
while 1 of 3 my rival did were, so I won.
2 of 3 poems I sent to a modern journal
were accepted,
while my rival had 3 of 5, so I won.
But overall, my rival had half of hers accepted
and I did not,
so she won
after all. I was
confounded! I found
that numbers don't lie,
but don't explain
why. Why
try comparing if comparison can be
reversed with a Peterson roll
by underdog wrestling
data, rival, or self?
When my parts are
summed,
am I less than some
of my parts?

„Verwirrt“

3 von 8 Gedichten
die ich einreichte
bei einer Klassik-Zeitschrift
wurden angenommen,
und 1 von 3 von meiner Rivalin, also gewann ich.
2 von 3 Gedichten an eine moderne Zeitschrift
wurden angenommen,
und von meiner Rivalin 3 von 5, also gewann ich.
Doch insgesamt wurde von ihr die Hälfte angenommen
und von mir nicht,
also gewann sie
am Ende doch. Ich war
verwirrt! Ich fand
dass Zahlen nicht lügen,
aber auch nicht erklären
warum. Warum
versucht man zu vergleichen, wenn Vergleiche
umgekehrt werden können, wie ein geschickter Griff
eines Außenseiters beim Ringen
von Daten, der Rivalin, von mir selbst?
Wenn meine Teile
zusammengefasst werden,
bin ich dann weniger als
die Summe/einige meiner Teile?

- Veranschauliche diese Umkehrung in einer Darstellung als Körper oder Fläche (für Anregungen siehe Lesser, 2001).
- Recherchiere den englischen Titel des Gedichts, „confounded“, und finde heraus, was er im Alltag und in der Statistik bedeutet. Erläutere, warum der Verfasser gerade diesen Begriff gewählt hat, um Verwunderung zu beschreiben.
- Im englischen Gedicht wird die „Peterson roll“ aus dem Ringsport erwähnt. Finde heraus, worum es sich dabei handelt.

Als weiterführende herausfordernde Aufgabe: Die Gesamtzahl der eingereichten Gedichte (also die Summe der vier Nenner) ist 19. Kannst du die Anzahlen der Gedichte so verändern, dass die Umkehrung weiterhin gilt, aber für weniger als 19 Gedichte insgesamt? (Spoiler: Diese Frage wird als Problem 5321 unter <http://ssma.play-cello.com/wp-content/uploads/2016/03/Feb-2015.pdf> beantwortet.)

7 Diskussion

Statistische Ideen auf fesselnde und für eine breite Öffentlichkeit zugängliche Weise darzustellen, ist eine wichtige Fähigkeit für Lehrkräfte und Forscher*innen. Das Simpson-Paradoxon stellt ein Instrument dar, mit dem die Relevanz des Verständnisses multivariaten Denkens adressiert werden kann. Wir haben das Simpson-Paradoxon genutzt, um mit einem breiten Spektrum von amüsanten Elementen aufzuzeigen, wie lehrreiche Songs, Cartoons, Gedichte etc. effizient und einprägsam eingesetzt werden können, um in ein statistisches Thema einzuführen oder es zu näher zu beleuchten. Wir freuen uns darauf, auch in Zukunft weitere Ideen und Beispiele zu teilen. Wenn Sie Kommentare zu dieser Kolumne oder Vorschläge für die Sammlung auf CAUSEweb.org haben, können Sie uns gerne per E-Mail (dkp13@psu.edu) kontaktieren.

Danksagung

Die Arbeit wurde unterstützt durch das Projekt SMILES, NSF/EHR/DUE 1544426 (PSU), 1544237 (UTEP). Alle Meinungsäußerungen, Ergebnisse und Schlussfolgerungen in diesem Material sind den Autoren zuzuordnen und stellen nicht notwendigerweise die Sichtweise der National Science Foundation dar.

Literatur

Champkin, J. (2011). Eveline Pye: Poetry in numbers. *Significance*, 8(3), 127–130.

- Friedlander R.; Wagon, S. (1993). Double Simpson's paradox. *Mathematics Magazine*, 66(4), 268.
- Gou, J.; Zhang, F. (2017). Experience Simpson's paradox in the classroom. *The American Statistician*, 71(1), 61–66.
- Klein, G.; Dabney, A. (2013). *The cartoon introduction to statistics*. New York: Hill and Wang.
- Lesser, L. (2018). Classroom notes: One in ten. *Teaching Statistics*, 40(1), 33–34.
- Lesser, L. M.; Wall, A. A.; Carver, R. H.; Pearl, D. K.; Martin, N.; Kuiper, S., et al. (2013). Using fun in the statistics classroom: An exploratory study of college instructors' hesitations and motivations. *Journal of Statistics Education*, 21(1), 1–33.
- Lesser, L. (2010). Confounded. *Mathematical Intelligencer*, 32(4), 53.
- Lesser, L. (2001). Representations of reversal: An exploration of Simpson's paradox. In A. A. Cuoco & F. R. Curcio (Eds.), *The roles of representation in school mathematics* (pp. 129–145). Reston, VA: NCTM.
- Neumann, D. L.; Hood, M.; Neumann, M. M. (2009). Statistics? You must be joking: The application and evaluation of humor when teaching statistics. *Journal of Statistics Education*, 17:2.
- Orsman, H.; Moore, J. (1988). *Heinemann dictionary of New Zealand quotations*. Auckland: Heinemann.
- Özdoğan, A. A.; McMorris, R. F. (2013). Humorous cartoons in college textbooks: Student perceptions and learning. *Humor*, 26(1), 135–154.
- Schild, M. (2006). Understanding confounding from lurking variables using graphs. *Stats: The Magazine for Students of Statistics*, 46, 14–18.
- Schild, M. (1999). Simpson's paradox and Cornfield's conditions. *Proceedings of the Joint Statistical Meeting JSM, Section on Statistical Education*, 106–111.
- Schneiter, K. (2013). An applet for the investigation of Simpson's paradox. *Journal of Statistics Education*, 21(1), 1–20.
- Sormani, M. P. (2009). The Will Rogers phenomenon: The effect of different diagnostic criteria. *Journal of the Neurological Sciences*, 287, S46–S49.
- Trumpower, D. (2010). Mad Libs statistics: A 'happy' activity. *Teaching Statistics*, 32(1), 17–20.

Anschriften der Verfasser

Lawrence M. Lesser
Mathematical Sciences Department
The University of Texas at El Paso
El Paso, Texas, USA
lesser@utep.edu

Dennis K. Pearl
Department of Statistics
Pennsylvania State University
University Park, Pennsylvania, USA
dkp13@psu.edu