mehr Spielen. Je nach Leistungsstärke können Informationen zur Lösung in die Lerngruppe hineingegeben werden.

10 Der Werkzeugeinsatz

Bei dieser Aufgabe und der kognitiven Zielsetzung der Unterrichtseinheit empfiehlt es sich, bewusst auf einen Taschenrechner zu verzichten, diese Bedingung sollte evtl. konkret in der Aufgabenstellung aufgenommen werden. Falls in der Lerngruppe die Notwendigkeit einer mathematischen Untersuchung wegen des fehlenden kognitiven Konflikts nicht erkannt wird, kann man die Situation mit einem kleinen Computerprogramm simulieren. Dazu müsste man natürlich wissen, wie man die unterschiedlichen Siegwahrscheinlichkeiten im Rechner simulieren kann. Man arbeitet mit Zufallszahlen modulo 3.

11 Abschließende Bemerkung

Das Ergebnis der Aufgabe steht nicht so sehr im Zentrum des Arbeitens. Entscheidender sind hier die prozessbezogenen Kompetenzen, die Lösungsstrategien, die hier gefördert werden und etwas abstrahiert tragfähig für weiteres mathematisches Arbeiten sind. Daher ist es essenziell, diese im Unterricht konkret herauszuarbeiten.

Literatur

Althoff, Heinz (1985 a): *Wahrscheinlichkeitsrechnung und Statistik*. Stuttgart: Metzler.

Althoff, Heinz (1985 b): *Wahrscheinlichkeitsrechnung und Statistik, Lösungen*. Stuttgart: Metzler Schulbuchverlag.

Bruss, F. Thomas (2008): *Tipps für Koch, Ypsilanti & Co*. In: Die Welt-Online, URL: www.welt.de/wissenschaft/article1720834/tipps_fuer_Koch_Ypsilanti_amp_Co.html. (Stand 25.3.2009)

Büchter, Andreas; Henn, Hans-Wolfgang (2006): *Elementare Stochastik.* 2. überarbeitete und erweiterte Auflage. Berlin/Heidelberg: Springer.

Engel, Arthur (1973): *Wahrscheinlichkeitsrechung und Statistik. Band 1*. Stuttgart: Klett.

Ministerium für Schule und Weiterbildung NRW (2007): *Kernlehrplan Mathematik für die Sekundarstufe I des Gymnasiums.* Fechen: Ritterbach.

Tietze, Uwe-Peter; Klika, Manfred; Wolpers, Hans (2002): *Mathematikunterricht in der Sekundarstufe II, Band 3.* Braunschweig/Wiesbaden: Vieweg.

Anschrift des Verfassers

Martin Epkenhans
Fakultät für Elektrotechnik, Informatik und Mathematik
Universität Paderborn
33098 Paderborn
martine@uni-paderborn.de

Ein kleines Simpson-Paradoxon bei den Ergebnissen von PISA-E

Renate Motzer, Augsburg

Zusammenfassung: *Vergleicht man die Ergebnisse der Leistungen bayerischer und sächsischer Schüler bei PISA-E 2006, so findet man zwar meist nur hauchdünne Unterschiede, aber diese widersprechen sich interessanterweise entsprechend des Simpson-Paradoxons.*

Im nationalen Vergleich haben bei PISA 2006 bekanntlich die sächsischen Schülerinnen und Schüler (der Einfachheit wegen im Weiteren unter „Schüler" zusammengefasst) am besten abgeschnitten und liegen diesmal vor den Jugendlichen aus Bayern.

Am deutlichsten ist ihr Vorsprung im Bereich der Naturwissenschaften (Sachsen: 541 Punkte, Bayern 533). In Mathematik (Sachsen 523, Bayern 522) und im Leseverständnis (Sachsen 542, Bayern 541) ist der Vorsprung nur hauchdünn (und natürlich in keiner Weise signifikant).

Schaut man sich aber z. B. nur die Leistungen von Gymnasiasten an, so ist bei den Naturwissenschaften Bayern mit den Sachsen gleichauf (619), in Mathematik (Bayern 608, Sachsen 599) und im Lesen (Bayern 598, Sachsen 587) erzielten die bayerischen Schülerinnen und Schüler einen kleinen Vorsprung.

Paradoxerweise kann nun festgestellt werden, dass es in Sachsen auch bei den Nichtgymnasiasten in Mathematik und bei den Leseleistungen keinen Vorsprung gegenüber Bayern gibt (Mathematik: Sachsen 487, Bayern 489, Lesen: Sachsen 477, Bayern 478). Nur in den Naturwissenschaften sind die sächsischen Nichtgymnasiasten mit 504 zu 500 Punkten ein bisschen erfolgreicher, aber auch hier ist der Unterschied nur halb so groß wie der, der sich ergibt, wenn man alle Schülerinnen und Schüler zusammennimmt. Da in Sachsen nicht zwischen Haupt- und Realschule getrennt wird, kann in der Gruppe der Nichtgymnasi-

asten nicht weiter differenziert werden. Nochmal alle Zahlen im Überblick:

	NW	g	ng	L	g	ng	M	g	ng
B.	533	619	500	541	598	478	522	608	489
S.	541	619	504	542	587	477	523	599	487

Die Unterschiede sind zwar nur hauchdünn, aber ein bisschen verblüffen kann es einen trotzdem: Da sind die bayerischen Schülerinnen und Schüler in Mathematik und im Lesen sowohl bei den Gymnasiasten als auch bei den Nichtgymnasiasten besser als die sächsischen, in der Gesamtwertung aber liegt jeweils Sachsen vorne. Dies scheint der Logik zu widersprechen.

Man stößt hier auf einen Fall des sog. Simpson-Paradoxons, das sich erst verstehen lässt, wenn man die Quoten berücksichtigt, nämlich wie viel Prozent eines Jahrgangs auf das Gymnasium gehen. In Bayern sind das nur 27,5 %, in Sachsen immerhin 32 %.
Dieser Unterschied von 4,5 % spielt eine Rolle. Wenn man die schwächsten 16 % der sächsischen Gymnasiasten (welche 4,5 % aller sächsischen Schüler ausmachen), aus dem Gymnasium weg in andere Schulen schicken würden, so kämen sowohl die sächsischen Gymnasiasten als auch die sächsischen Nichtgymnasiasten auf bessere Schnitte, denn unter den Nichtgymnasiasten würden diese Schüler ja vermutlich wieder zu den besseren zählen.

Erst wenn man weiß, welcher Anteil der untersuchten Jugendlichen auf Gymnasien geht, kann man aus den nach Schulart getrennten Ergebnissen auf den Schnitt aller Schüler schließen. Das Gesamtmittel ist nämlich ein gewichtetes Mittel der jeweiligen Gruppenmittel.

Für Mathematik gilt:

27,5 % * 608 + 72,5 * 489 = 522 (gerundet, da auch die anderen Werte auf ganze Punktzahlen gerundet angegeben wurden)

Bayr. Gymnasialanteil * Bayr. Matheleistung Gymnasium + Bayr. nichtgymn. Anteil * Bayr. Matheleistung Nichtgymn. = Bayerisches Mittel

und

32 % * 599 + 68 % * 487 = 523 (gerundet)

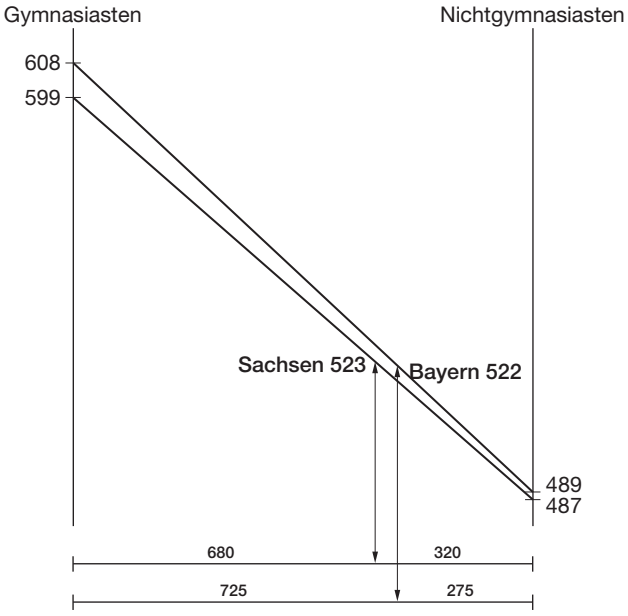
Sächs. Gymnasialanteil * Sächs. Matheleistung Gymnasium + Sächs. nichtgymn. Anteil * Sächs. Matheleistung Nichtgymn. = Sächsisches Mittel

(De facto habe ich umgekehrt aus den Durchschnittspunktzahlen für alle Schüler und aus denen für die Gymnasiasten aufgrund dieses Zusammenhangs auf die Leistungen der Nichtgymnasiasten zurückgerechnet.)

Statt mit Prozentzahlen zu rechnen, könnte man auch natürliche Häufigkeiten einführen, da diese von Schülern oft besser verstanden werden. Allerdings muss ich hier einen fiktiven Wert (z. B. je 1 000 bayerische und 1 000 sächsische Schüler) wählen, da mir die exakten Daten nicht bekannt sind.
Von 1 000 bayerischen Schülern sind 275 auf einem Gymnasium. Diese erzielen zusammen 275 * 608 Punkte. Die 725 Jugendlichen, die nicht auf ein Gymnasium gehen, erzielen miteinander 725 * 489 Punkte. Insgesamt haben es 1 000 bayerische Schüler damit auf 521 725 Punkte gebracht.

Bei den sächsischen Schülern schaut es so aus:
Die 320 Gymnasiasten erhalten zusammen 320 * 599 Punkte, die restlichen 680 schaffen 680 * 487 Punkte. Zusammen sind dies 522 840 Punkte für 1 000 sächsische Schüler.

Folgende Grafik (Idee nach Tan 1986) kann diesen Zusammenhang ebenfalls verdeutlichen:



Man beachte bei dieser Grafik, dass die Häufigkeiten der Gymnasiasten auf den unteren Linien von rechts nach links dargestellt wurden, obwohl die Ergebnisse der Gymnasiasten auf der linken Skala eingetragen wurden. Dies ist nötig, denn sieht man sich die Geradengleichungen an, so gilt etwa für die bayerischen Schüler für die Berechnung des Gesamtschnittes (mit x als relativer Häufigkeit der Gymnasiasten):
$x * 608 + (1 - x) * 489 = x * 119 + 489$,
es ist also eine Gerade, die beim rechten Skalenwert 489 beginnt und die x-Richtung ungewöhnlicherweise nach links verläuft.

Eine weitere mögliche Fragestellung wäre: Wie viel Prozent der bayerischen Schüler dürfen maximal aufs

Gymnasium gehen, damit die bayerischen Schüler sowohl bei den Gymnasiasten als auch bei den Nichtgymnasiasten besser dastehen als die sächsischen?

Gehen wir dabei davon aus, dass sich die Leistungen der bayerischen Schüler insgesamt nicht verbessern würden, auch wenn ein paar mehr aufs Gymnasium gingen. (Würden sich die Leistungen doch verbessern, was man von Anspruch des Gymnasiums her vermutlich erwarten sollte, dürfte der sächsische Schnitt ja auch insgesamt überboten werden.)
Um besser zu sein als die sächsischen Gymnasiasten, müsste der bayerische Schnitt bei mindestens 600 Punkten liegen, bei den Nichtgymnasiasten mindestens bei 488 Punkten.

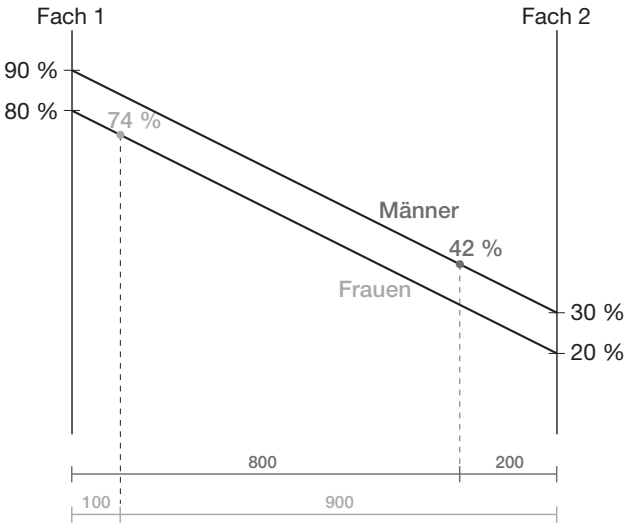
Es gilt also
$x * 600 + (1 - x) * 488 = 522$
zu lösen und man erhält: $x = 30{,}4\ \%$.

Auf Anregung von Joachim Engel sei weiterhin angemerkt: Man könnte auch ein Szenario konstruieren, das einen zum Schluss kommen lässt, dass die Schüler in Nordrhein-Westfalen bessere Matheleistungen aufweisen als die Schüler in Bayern und Sachsen.
Wir nehmen das beste Prozent aller NRW-Schüler und stecken sie in ein virtuelles Gymnasium Gstar. Die übrigen 99 % kommen in die Restschule Reststar.
Klar, dass der Durchschnitt in Gstar in NRW besser ist als die bayerischen Gymnasiasten. Es ist ja die höchste NRW-Elite. Die restlichen 99 % vom Reststar sind auch nicht so schlecht, naja etwas schlechter als der NRW-Durchschnitt (493). Das könnte, da ja nur 1% fehlt, mehr als 489 sein, der Schnitt der bayerischen Nichtgymnasiasten. Bremen hat leider auch für solch einen Kniff keine Chance, da der Gesamtschnitt mit 478 dort unter dem der bayerischen Nichtgymnasiasten liegt.
Aber das ist nur eine Konstruktion. Das Schöne an den echten PISA-Daten ist, dass sie ein aktuelles reales Beispiel liefern, das man auch im Unterricht verwenden kann, wenn das Simpson-Paradoxon besprochen wird. Sie zeigen, dass dieser Effekt tatsächlich auch in unserem Umfeld auftreten kann.

Bei fiktiven Daten kann man freilich leichter rechnen und deutlichere Effekte erzielen (vgl. z. B. Henze 2006, Meyer 1995, Getrost u. Stein 1994).
Fiktive Daten können auch Anlass sein, selbst Aufgaben zu konstruieren bzw. Schüler solche Aufgaben konstruieren zu lassen.
Die Vorgaben könnten ähnlich wie oben sein: die Werte alle Teilgruppen beider Bereiche und die Zusammensetzung für einen Bereich. Dann kann gefragt werden, wie sich der andere Bereich zusammensetzen kann, dass sich eine paradoxe Situation ergibt. Wie im obigen Beispiel kann die Prozentzahl x ausgerechnet werden (statt der Gleichung wäre dann eine lineare Ungleichung zu lösen).
Die Lösung kann auch grafisch am Diagramm ermittelt werden (wenn es maßstabsgetreu gezeichnet wird). Die Frage ist, wann der Punkt der unteren Gerade über dem gewichteten Mittelwert liegt, der auf der oberen Gerade eingezeichnet wird.
An der Veranschaulichung ist auch ersichtlich, dass sich jeder Punkt auf den die linke und rechte Seite verbindenden Strecken durch eine geeignete Verteilung erreichen lässt. Es muss also nur der Endpunkt der oberen Strecke unterhalb des Anfangspunktes der unteren Strecke liegen und schon kann ein Simpson-Paradoxon entstehen.

Passend etwa zum Beispiel aus Henze (2006, S. 109), schauen die Tabelle und die zugehörige Veranschaulichung so aus:

	Frauen		Männer	
	insges.	zugel.	insges.	zugel.
Fach 1	900	720	200	180
Fach 2	100	20	800	240
Summe	1 000	740	1 000	420



An diesem Bild sieht man sehr gut, dass es einige Möglichkeiten für ein Simpson-Paradoxon gibt.
Legt man die Punkte auf den Strecken anders, so kann dafür wiederum die Tabelle der absoluten Häufigkeiten errechnet werden. Diese Umkehrung der Fragestellung dürfte ein vertieftes Verständnis des Phänomens fördern.

Literatur

http://pisa.ipn.uni-kiel.de/Zusfsg_PISA2006_national.pdf

Profil – Das Magazin für Gymnasium und Gesellschaft, 12/2008, S. 6 ff.

Henze, N. ([6]2006): *Stochastik für Einsteiger. Wiesbaden: Vieweg.*

Meyer, J. (1995): *Einfache Paradoxien der beschreibenden Statistik. In: Stochastik in der Schule 15(2), 27–50.*

Getrost, G. (1994); Stein, G: *Fehler und Fallen der Stochastik im Stochastikunterricht. In: Prax. Math 36(6), 249–253.*

Tan, A. (1986): *A Geometric interpretation of Simpsons Paradoxon. In: Call. Math. Jour. 17 (4),* S. 340 f.

Anschrift des Verfassers

Renate Motzer
Didaktik der Mathematik
Universität Augsburg
Universitätsstr. 10
86135 Augsburg
Renate.Motzer@math.uni-augsburg.de

Polyalphabetische Verschlüsselung und Stochastik

Thorsten Mehlich, Bochum

***Zusammenfassung**: Am Beispiel der Vigenère-Verschlüsselung sollen die Themen relative Häufigkeiten und Laplace-Verteilungen verdeutlicht werden. Dazu wird als erstes die Vigenère-Verschlüsselung vorgestellt. Anschließend wird der Kasiski-Test beschrieben, um ein Gefühl zu vermitteln, wie relative Häufigkeiten hier eine Rolle spielen. Danach folgt eine genaue stochastische Betrachtung der Verschlüsselung und die Herleitung des Friedman-Koinzidenzindexes.*

1 Die Vigenère-Verschlüsselung

Definition 1.1 *Eine unverschlüsselte Nachricht nennt man* **Klartext**.

Definition 1.2 *Verschlüsselt man eine Nachricht, so nennt man das Ergebnis* **Kryptogramm** *oder* **Kryptotext**.

Definition 1.3 *Bei einem* **Substitutionsalphabet** *wird jeder Klartextbuchstabe durch genau einen Buchstaben des Alphabets ersetzt.*

Definition 1.4 *Im Fall einer* **polyalphabetischen Verschlüsselung** *kann ein Klartextbuchstabe durch mehr als nur einen Buchstaben ersetzt werden. Wie er substituiert wird, hängt von der Verschlüsselungsmethode ab.*

Bei der Vigenère-Verschlüsselung handelt es sich um eine polyalphabetische Verschlüsselungsmethode. Hierbei benutzt man für verschiedene Buchstaben des Klartextes unterschiedliche Substitutionsalphabete. Ein gutes Konzept hierfür machte bereits Vigenère im 16. Jahrhundert. Er benutzte dazu die nach ihm benannte Vigenère-Tabelle (Tab. 1).

Zur Verschlüsselung benötigt man ein Schlüsselwort, das periodisch wiederholt wird. Sei das Schlüsselwort z. B. „BACH“ und die Nachricht „GEHENACHWESTEN“.

Schlüssel:	B	A	C	H	B	A	C	H
Klartext:	G	E	H	E	N	A	C	H
Kryptogramm:	H	E	J	L	O	A	E	O
Schlüssel:	B	A	C	H	B	A		
Klartext:	W	E	S	T	E	N		
Kryptogramm:	X	E	U	A	F	N		

Jeden einzelnen Buchstaben des Klartextes sucht man in der obersten Zeile (siehe Tab. 1). Dann geht man in der entsprechenden Spalte so weit nach unten, bis man die Zeile mit dem Buchstaben (linker Rand der Tabelle) des Schlüssels findet. Beim Schnittpunkt der Spalte und der Zeile steht der Buchstabe für das Kryptogramm. In dem Beispiel beginnt man mit dem *G* in der obersten Zeile und geht dann zwei Zeilen nach unten zum B-Alphabet. Beim Schnitt der Senkrechten und der Waagerechten findet man das *H* für die verschlüsselte Nachricht. Um eine Nachricht zu dechiffrieren, geht man genau umgekehrt vor.

1.1 Angriff auf eine Vigenère-Verschlüsselung

Wenn ein Angreifer eine auf diese Art und Weise verschlüsselte Nachricht dechiffrieren möchte, dann