Was sagen signifikante Ergebnisse? - Zu einem Beispiel aus der Zeitung

Bernhard Waldmüller, Espelkamp

Zusammenfassung: Unter dem Titel "Glaubenskrieg um Krümmel" beschreibt Gerd Rosenkranz in einem sehr lesenswerten Zeitungsartikel (Rosenkranz 1998) den Stand der Auseinandersetzung um die Frage, ob Atomanlagen Leukämie bei Kindern verursachen oder nicht. Dabei spielt eine wichtige Rolle, ob Daten signifikante Häufungen aufweisen und welche Schlüsse man daraus ableiten kann. Von diesem Artikel habe ich mich anregen lassen und ein Fallbeispiel konstruiert, an dem ich mit Schülern arbeiten kann: signifikante Häufungen feststellen, sehen, wie damit argumentiert wird, erfahren, daß statistische Aussagen letztlich keine abschließenden Antworten geben.

1. Die Daten

Atomkraftwerke stehen in dem Verdacht, bei Kindern Leukämie zu verursachen. Deshalb wird über Leukämiefälle in der Umgebung von Atomkraftwerken genau Buch geführt. Manche betrachten dabei einen Kreis von 5 *km* um das Kraftwerk, andere einen Kreis von 15 *km*.

In einem Zeitraum von einigen Jahren wurden bei den 15.000 Einwohnern des 5 *km*-Kreises um ein Kernkraftwerk neun Fälle von Leukämie bei Kindern und Jugendlichen beobachtet - im Bundesdurchschnitt ist es nur ein Fall auf 10.000 Einwohner! Insgesamt 310.000 Einwohner leben mehr als 5 *km*, aber höchstens 15 *km* vom Kraftwerk entfernt. 110.000 von ihnen gehören zum gleichen Landkreis wie die Leute im inneren Bereich, 200.000 gehören zu einer großen Stadt, deren Gebiet in den 15 *km*-Bereich hineinragt. Bei diesen 310.000 Leuten entspricht der Anteil der Leukämiefälle genau dem Bundesdurchschnitt.

2. Die Aufgaben

1. Überprüfe und kommentiere die folgenden Aussagen.

a) Kritische Wissenschaftler schlagen Alarm: Im 5 *km*-Kreis ist die Anzahl der Leukämiefälle dramatisch erhöht.

b) Das Gesundheitsamt des Landkreises beruhigt: Im Kreisgebiet ist zwar eine gewisse Erhöhung festzustellen, aber es liegt keine signifikante Abweichung vom Bundesdurchschnitt vor.

c) Das Bundesumweltministerium gibt bekannt, daß im Umkreis von 15 *km* um die kerntechnische Anlage kein erhöhtes Leukämierisiko für Kinder feststellbar ist.

2. Ein atomkritischer Wissenschaftler spricht empört von "Verdünnung kritischer Daten durch statistische Tricks". Was ist damit gemeint?
3. In einem kleinen Dorf, genau 16 *km* vom Kraftwerk entfernt, erkrankt ein Kind an Leukämie. Der Sprecher einer Bürgerinitiative macht dafür das Kraftwerk verantwortlich. Sein Argument: "Die Wahrscheinlichkeit, daß bei unseren 200 Einwohnern ein oder sogar mehr Fälle von Leukämie auftreten, liegt unter 2 %! Nach der Statistik hätte deshalb im Dorf kein Krankheitsfall auftreten dürfen." Hat er richtig gerechnet?
4. Welche Schlüsse würdest *du* aus den Daten ziehen? Was schlägst du vor?

3. Überlegungen zur Lösung

Zu Aufgabe 1

Wir nehmen an, daß die Zufallsgröße S_n: "Anzahl der Leukämiefälle im Untersuchungszeitraum unter *n* Bewohnern" binomial-verteilt ist mit der Erfolgswahrscheinlichkeit $p = 1/10.000$. Dann ist

$$\mu_n := \mathrm{E}\,(S_n) = n\,p\ ,\ \mathrm{V}\,(S_n) = n\,p\,q \quad \text{und} \quad \sigma_n := \sqrt{n\,p\,q}$$

Unter den 15.000 Bewohnern des 5 *km*-Kreises müßte man $\mu_{15.000} = 1{,}5$ Fälle erwarten. Es ist

$$\frac{9-1{,}5}{\sigma_{15.000}} \approx 6{,}124$$

der beobachtete Wert liegt um mehr als das 6-fache der Standardabweichung $\sigma_{15.000}$ vom Erwartungswert weg. Es handelt sich um eine hoch signifikante Abweichung, die Zahl der Fälle im 5 *km*-Kreis ist tatsächlich dramatisch erhöht. Zum Landkreis gehören außer den 15.000 Bewohnern des 5 *km*-Kreises weitere 110.000 Menschen. Die Anzahl der Leukämiefälle beträgt $9 + 110.000\,p = 9+11 = 20$. Die Abweichung dieses Wertes vom Erwartungswert $125.000\,p = 12{,}5$ beträgt

$$\frac{20-12{,}5}{\sigma_{125..000}} \approx 2{,}121$$

Standardabweichungen $\sigma_{125.000}$. Die Aussage des Kreisgesundheitsamtes ist von daher problematisch, aber noch möglich.

Im 15 *km*-Kreis leben insgesamt 15.000 + 310.000 Menschen, und es wurden

$$9 + 310.000p = 40$$

Leukämiefälle beobachtet. Die Abweichung vom Erwartungswert $325.000p = 32{,}5$ beträgt $7{,}5 / \sigma_{325.000} \approx 1{,}316$ Standardabweichungen $\sigma_{325.000}$. Die Aussage des Ministeriums kann nicht beanstandet werden.

Zu Aufgabe 2

Was in dieser Aufgabe mit "Verdünnung kritischer Daten" gemeint ist, ist klar. Im 5 *km*-Kreis erhalten wir pro Einwohner 9/15.000 = 0,0006 Fälle, im Landkreis 20/125.000 = 0,00016 Fälle, im 15 *km*-Kreis 40 / 325.000 = 0,000123 Fälle.

Allgemein gilt: Wenn man zu den 15.000 Einwohnern des 5 *km*-Kreises x Einwohner aus der Umgebung, also aus einer Grundgesamtheit mit dem Risiko $p = 0{,}0001$, hinzufügt, ist die Häufigkeit der Leukämiefälle gegeben durch

$$\mathrm{h}(x) := \frac{9 + p\ x}{15.000 + x}$$

Es ist h(0) = 9 / 15.000 = 0,0006, h(x) ist für $x \geq 0$ streng monoton fallend und strebt gegen $p = 0{,}0001$ für $x \to \infty$. Man kann also im Prinzip durch geeignete Wahl von x jede Häufigkeit zwischen 0,0006 und p einstellen. Das ist natürlich problematisch.

Zu Aufgabe 3

Die Frage in dieser Aufgabe beleuchtet die Sache von einer anderen Seite her. Die Wahrscheinlichkeit, daß im Dorf im Beobachtungszeitraum ein Leukämiefall oder sogar noch mehr auftreten, ist

$$P(S_{200} \geq 1) = 1 - P(S_{200} = 0) = 1 - 0{,}9999^{200} \approx 0{,}0198\,,$$

und das sind tatsächlich nur knapp 2%. (Hätte sich der Sprecher der Bürgerinitiative nur auf die Familie des kranken Kindes oder gar nur auf das Kind selbst bezogen, wäre das Ergebnis noch deutlicher ausgefallen.) Jedenfalls ist die Rechnung der Bürgerinitiative fehlerfrei.

Zu Aufgabe 4

Die Frage in dieser Aufgabe soll die Schüler nicht ermutigen, mit ihren Kenntnissen auf komplizierte Fragen abschließende Antworten zu geben. Aber ihre Beschäftigung mit Stochastik sollte sie in die Lage versetzen, den Zeitungsartikel mit Gewinn lesen zu können. Welche Schlüsse kann man aus den vorliegenden Daten ziehen? Offensichtlich sind die Fallzahlen im 5 *km*-Kreis um das Kraftwerk deutlich erhöht. Ob das Kraftwerk für diese Erhöhung verantwortlich ist, kann nur durch eine inhaltliche Untersuchung geklärt werden. Auf welchem Wege könnte das Kraftwerk auf die Umgebung eingewirkt haben? Wenn dies durch die Luft geschehen ist, sollte die in dem Gebiet vorherrschende Windrichtung von Einfluß sein. Welche Wege nimmt das Abwasser des Werks, auf welchen Routen finden Materialtransporte statt? Wo wohnen die Mitarbeiter - falls diese Radioaktivität nach draußen tragen sollten? Wie weit könnte direkte Strahlung reichen? Im Zuge der Untersuchung solcher Fragen würden Hypothesen formuliert, die wieder mit statistischen Methoden überprüft werden müßten.

Zum Grad der Auffälligkeit der vorliegenden Daten können wir mit unseren Mitteln durchaus Näheres sagen. Da $V(S_{125.000}) = 125.000\,p\,(1-p) \approx 12{,}5$ groß genug ist, können wir mit Hilfe der Gaußfunktion die Wahrscheinlichkeit $P(S_{125.000} \geq 20)$ näherungsweise berechnen, daß die beobachtete Fallzahl 20 oder eine noch höhere unter 125.000 zufällig gewählten Bundesbürgern auftritt. Es ist

$$P(S_{125.000} \geq 20) \approx 1 - \Phi\left(\frac{19{,}5 - 12{,}5}{\sigma_{125.000}}\right) \approx 1 - \Phi(1{,}980) \approx 1 - 0{,}9761 = 0{,}0239.$$

Um die Bedeutung dieses Werts einschätzen zu können, denken wir uns die N=80.000.000 Bundesbürger in 80.000.000 / 125.000 = 640 Gruppen zu 125.000 eingeteilt. Wir bilden zu jeder Gruppe eine Zufallsgröße, die den Wert 1 annimmt, falls es in der Gruppe 20 oder mehr Leukämiefalle gibt, und sonst den Wert 0. Wenn wir nur eine der Gruppen betrachten, können wir ihre Mitglieder als zufällig gewählt ansehen, ihre Zufallsgröße nimmt den Wert 1 also mit der eben berechneten Wahrscheinlichkeit 0,0239 an. Die Gesamtzahl der auffälligen Gruppen ist durch die Summe der Zufallsgrößen gegeben. Nach der Regel für den Erwartungswert einer Summe von Zufallsgrößen ist der Erwartungswert der Gesamtzahl dann $640 \times 0{,}0239 \approx 15{,}3$. Man muß also durchaus damit rechnen, daß es in der Bundesrepublik eine ganze Reihe von Gruppen zu 125.000 Personen mit mindestens 20 Leukämiefällen gibt, auch ohne daß eine aufweisbare Ursache dahinter steckt.

Die entsprechende Rechnung ergibt für den 15 *km*-Kreis

$$P(S_{325.000} \geq 40) \approx 1 - \Phi\left(\frac{39{,}5 - 32{,}5}{\sigma_{325.000}}\right) \approx 1 - \Phi(1{,}228) \approx 1 - 0{,}8903 = 0{,}1097,$$

$$0{,}1097 \times \frac{N}{325.000} \approx 27$$

Leider läßt sich die Gaußfunktion beim 5 *km*-Kreis nicht anwenden, weil die Varianz $V(S_{15.000}) \approx 1{,}5$ zu klein ist. Mit Hilfe der Poisson-Näherung erhält man

$$P(S_{15.000} \geq 9) \approx 1 - \sum_{k=0}^{8} \frac{1{,}5^k}{k!} \; e^{-1{,}5} \approx 0{,}0000277\,.$$

Nun ist $0{,}0000277 \times N / 15.000$ nur ungefähr 0,148, das unterstreicht, daß die Fallzahl im 5 *km*-Kreis wirklich außergewöhnlich hoch ist.

Die Daten des Dorfes ergeben

$$P(S_{200} \geq 1) \times \frac{N}{200} \approx 0{,}0198 \times 400.000 = 7920$$

zu erwartende Gruppen mit Leukämiefällen bei Einteilung der Bevölkerung in Gruppen zu 200 Personen. Obwohl die beobachtete Anzahl 1 um fast $7 \times \sigma_{200}$ vom Erwartungswert μ_{200} abweicht, würde ich dieser Beobachtung keine große Auffälligkeit bescheinigen.

Eine letzte Anmerkung zur Argumentation der Bürgerinitiative. Hauptgewinne im Lotto sind selten, aber es gibt welche. Wenn nun irgendwo ein Gewinner wohnt, kann jemand aus seiner Umgebung behaupten, daß hier etwas nicht mit rechten Dingen zugegangen sei, weil es in einer (genügend klein gewählten) Gruppe von rechts wegen keinen Lottogewinner geben dürfte. Ein Leukämiefall ist eine sehr ernste Angelegenheit, und die Betroffenen haben ein Recht darauf, daß man ihren Fragen nachgeht. Aber von der Statistik her besteht meines Erachtens wenig Hoffnung, daß bei einer Untersuchung der Situation des Dorfes etwas herauskommt. Ich würde dafür plädieren, die Untersuchung auf den 5 *km*-Kreis zu konzentrieren.

Literatur

Rosenkranz, Gerd: Glaubenskrieg um Krümmel - Die Frage, ob Atomanlagen bei Kindern Leukämie verursachen oder nicht, spaltet die Zunft der Epidemiologen. "Aus Wissenschaft und Technik" in der *Süddeutschen Zeitung* vom 26.3.98.

Bernhard Waldmüller, Samlandweg 19, 32339 Espelkamp