Stochastik-Abiturprüfung 2021 Baden-Württemberg – zwei denkwürdige Aufgaben zu Binomialtests

NORBERT HENZE, KARLSRUHE

Zusammenfassung: *Die in diesem Jahr im Wahlteil Stochastik der Hauptprüfung Abiturprüfung 2021 – Leistungsfach Baden-Württemberg gestellten Aufgaben zu Binomialtests stifteten unter Lehrkräften und vielen Prüflingen Verwirrung. Dieser Aufsatz beleuchtet die Gründe für diese Verwirrung, und er liefert ein weiteres Plädoyer dafür, das statistische Testen aus den Bildungsplänen zu verbannen und dadurch wieder mehr Mathematik zu ermöglichen.*

1 Einleitung

Vor einigen Monaten erhielt ich mehrere E-Mails von Lehrkräften und Fachberater(inne)n an Gymnasien in Baden-Württemberg. Den Anlass für diese überraschenden Kontaktaufnahmen zu einem Fachstochastiker an einer Universität bildeten zwei Aufgaben zur Abiturprüfung 2021 - Leistungsfach, die den Aufgabensatz 2 im Wahlteil Stochastik betrafen. Die Aufgaben (zu finden etwa unter der URL https://www.mathe-aufgaben.com) lauten:

Aufgabe C 2.1: Bei einem Glücksrad gibt es drei farbige Sektoren. Beim einmaligen Drehen beträgt die Wahrscheinlichkeit für rot 60%, für blau 30% und für grün 10%.

a) Das Glücksrad wird 20-mal gedreht. Bestimmen Sie für folgende Ereignisse jeweils die Wahrscheinlichkeit:
A: „Genau 15-mal erscheint rot.“
B: „Bei den ersten zehn Drehungen erscheint genau zweimal blau, insgesamt erscheint höchstens fünfmal blau.

b) Bei einem Glücksspiel darf man für einen Einsatz von 6 Euro das Glücksrad zweimal drehen. Wenn dabei genau einmal rot erscheint, dann erhält man einen bestimmten Auszahlungsbetrag. Wenn zweimal rot erscheint, dann erhält man das Siebenfache dieses Auszahlungsbetrags. Andernfalls erfolgt keine Auszahlung. Das Spiel ist fair. Bestimmen Sie den Auszahlungsbetrag für den Fall, dass genau zweimal rot erscheint.

c) Jemand vermutet, dass die Wahrscheinlichkeit für rot in Wirklichkeit geringer als 60% ist. Deshalb soll ein Hypothesentest durchgeführt werden. Dabei soll möglichst vermieden werden, dass irrtümlich von einer zu hohen Wahrscheinlichkeit für rot ausgegangen wird. Formulieren Sie eine Nullhypothese, die dieser Zielsetzung entspricht, und begründen Sie Ihre Wahl.

Aufgabe C 2.2: Bei einem Spielautomaten wird vermutet, dass die Gewinnwahrscheinlichkeit p größer als 10% ist. Die Vermutung wird mit Hilfe eines Hypothesentests mit dem Stichprobenumfang von $n = 200$ und einem Signifikanzniveau von 5% getestet. Als Nullhypothese wird $H_0 : p \geq 0{,}1$ gewählt. Formulieren Sie die Entscheidungsregel.

Tatsächlich gilt $p = 0{,}08$. Bestimmen Sie die Wahrscheinlichkeit für den Fehler zweiter Art.

Um die im Tenor mehr oder weniger gleichen Kritikpunkte der Lehrkräfte an diesen Aufgaben einordnen zu können, muss man wissen, dass ein in Baden-Württemberg weit verbreitetes Schulbuch der „Lambacher Schweizer“ ist (s. Dürr u.a. (2017)). In diesem Buch wird auf S. 278 als Faustregel formuliert, dass die zu bestätigende Aussage als Alternativhypothese H_1 zu formulieren sei, und diese Regel findet bei allen Aufgaben in diesem Buch Anwendung.

Hierauf bezieht sich auch der prototypische und von einer Lehrkraft mehr oder minder wörtlich übernommene Einwand:

> „Die in der Aufgabe C 2.1 c) zuerst genannte Vermutung würde zu einem skeptischen Glücksspieler passen, der das Glücksrad in Frage stellt. Die spätere Formulierung (Vermeidung hoher Wahrscheinlichkeiten) passt zu dem Glücksspielbetreiber, der vermeiden will, dass er zu oft Gewinne auszahlen muss. Das kann man doch nicht in EINER Aufgabe mit EINEM Test abfragen? In Dürr u.a. (2017) werden auf Seite 279 genau für solche Fragestellungen aus zwei Perspektiven zwei Nullhypothesen bzw. unterschiedliche Alternativhypothesen formuliert. Ich finde, diese Aufgabe ist nicht in Einklang mit dem Schulbuch, oder übersehe ich dabei etwas? Außerdem stört mich die Formulierung „Deshalb ... “ und „Dabei... “, weil hier zwei sich widersprechende Intentionen genannt werden. Hierdurch wird nicht gerade das Vertrauen in die Verlässlichkeit des Mathematik-Abiturs gefördert.
>
> Übrigens ist auch Aufgabe C 2.2 im gleichen Sinne widersprüchlich: Die Vermutung wird hier in H_0 gesteckt. Auch hier haben einige meiner Schüler die vor-

gegebene H_0 übersehen und ihre eigene formuliert, die natürlich „anders herum“ ist.“

2 Eine kritische Analyse

Der in der Testtheorie gängige Ansatz, die Wahrscheinlichkeit für einen Fehler erster Art, also eine Hypothese H_0 im Fall ihrer Gültigkeit abzulehnen, durch einen kleinen Wert wie z.B. 0,05 zu beschränken, zeigt, dass das Testen überhaupt nur dann sinnvoll ist, wenn der Fehler 1. Art und der Fehler 2. Art (letzterer entsteht, wenn eine Hypothese H_0 nicht gilt und man aufgrund des Testes keinen Widerspruch zu H_0 erhebt) unterschiedliche Konsequenzen beinhalten. Lehnt ein solcher sog. *Test zum Niveau* 0,05 die Hypothese H_0 nicht ab, so kann dieser Sachverhalt durchaus bedeuten, dass H_0 nicht gilt und nicht genügend Daten zur Verfügung standen, um einen signifikanten Widerspruch zu H_0 zu erhalten. Mit anderen Worten: Man kann bis auf Weiteres mit dem stochastischen Modell (bzw. den stochastischen Modellen) unter H_0 weiterarbeiten. Wird aber H_0 abgelehnt, so ist man sich „praktisch sicher“, dass H_0 nicht gilt, denn sonst wäre man aufgrund des Tests nur mit der kleinen Höchstwahrscheinlichkeit von 0,05 zur Ablehnung von H_0 gelangt. Damit ist aber klar, wie man die Alternative H_1 zu H_0 festlegt: Wenn man etwas vermutet, so möchte man ein Verfahren anwenden, dass einem *die Möglichkeit eröffnet, diese Vermutung signifikant abzusichern*. Genau das besagt die in Dürr u.a. (2017) auf Seite 279 formulierte Faustregel, die nur ein Grundprinzip beim statistischen Testen hervorhebt (s. z.B. Henze (2021), Kap. 30).

Versetzen wir uns jetzt einmal in die Situation eines Prüflings – nennen wir ihn Anja – der in der Stresssituation einer Abiturprüfung Aufgabe C 2.1 zu lösen versucht. Nachdem sich Anja mit Aufgabenteil b) beschäftigt hat, hat sie begriffen, dass das Auftreten von rot etwas Positives für den Spieler ist, denn in diesem Fall gewinnt er etwas. Mit diesem Gedanken im Hinterkopf liest Anja den ersten Satz von Aufgabenteil c). Ich weiß nicht, ob sich Anja fragt, wer der *jemand* ist, aber aus Sicht des Spielers sollte doch die Wahrscheinlichkeit für rot (mindestens) 0,6 sein. Das folgende Wort „Deshalb“ kann ja nur bedeuten, dass aufgrund der Vermutung des *jemand* ein Hypothesentest durchgeführt werden soll. Was sind dessen mögliche Fehler?

Testet man in einem ersten Szenario die Hypothese $H_0 : p \leq 0{,}6$ gegen die Alternative $H_1 : p > 0{,}6$, so ist der Fehler, den man tunlichst nicht begehen möchte, derjenige, dass in Wirklichkeit $p \leq 0{,}6$ ist, aber der Test aufgrund von zu vielen Erfolgen ernsthafte Zeifel an dieser Hypothese hegt und sie verwirft. Dieser Test kann *prinzipiell nur die Vermutung statistisch absichern, dass p größer als* 0,6 *ist*.

In einem zweiten Szenario seien die Hypothese $H_0 : p \geq 0{,}6$ und die Alternative $H_1 : p < 0{,}6$. In diesem Szenario möchte *jemand* möglichst vermeiden, dass der Test für den Fall, dass tatsächlich rot mit der (Mindest)Wahrscheinlichkeit 0,6 auftritt, der Test aufgrund zu weniger Erfolge zu einem signifikanten Widerspruch zu H_0 gelangt. Dieser Test ist also *von seinem Ansatz her in der Lage, die Vermutung statistisch abzusichern, dass p kleiner als* 0,6 *ist*.

In der Aufgabenstellung steht jetzt: „Dabei soll möglichst vermieden werden, dass irrtümlich von einer zu hohen Wahrscheinlichkeit für rot ausgegangen wird“. An dieser Stelle vermutet Anja wohl, dass Folgendes gemeint ist: In Wirklichkeit ist p höchstens gleich 0,6, und wir wollen vermeiden, dass ein Test „dieser Wirklichkeit“ – formuliert als Hypothese $H_0 : p \leq 0{,}6$ – aufgrund zu vieler Erfolge dazu führt, dass diese Hypothese abgelehnt wird (das Wort *ausgegangen* ist äußerst kritisch zu sehen). Natürlich kann man einen solchen Test durchführen, aber was sind dessen Konsequenzen? Wird die Hypothese nicht abgelehnt (hat man also kein signifikantes Ergebnis), so weiß man im Prinzip nichts. Vielleicht war die Anzahl der Erfolge (rot) nicht groß genug, um einen signifikanten Widerspruch zu H_0 herbeizuführen. Erhält man aber einen solchen Widerspruch, so kann man mit Fug und Recht behaupten, dass die Wahrscheinlichkeit für rot größer als 0,6 ist. Zu Begin von Teil c) wurde aber die Vermutung aufgestellt, dass die Wahrscheinlichkeit für rot *kleiner* ist als 0,6. Der Test im zweiten Szenario kann diese Vermutung **prinzipiell nicht signifikant absichern**. Laut Aufgabensteller(inne)n soll aber genau dieser Test durchgeführt werden.

Von schulischer Seite wurde mir aufgrund meiner Kritik entgegengehalten, dass es ein Irrtum sei, dass bei einem Hypothesentest eine „Vermutung“ grundsätzlich als Alternativhypothese zu wählen sei. Das sei aber keineswegs der Fall. Vielmehr entscheide erst die Zielsetzung des Testenden (eine solche Vermutung entweder zu untermauern oder aber zu entkräften) über die Festlegung der Nullhypothese.

Wir fragen uns: Soll Anja in der Stresssituation einer Abiturprüfung darüber spekulieren, wer den Test in Auftrag gibt? Sie muss, *weil jemand vermutet und deshalb* ein Hypothesentest durchgeführt wird, davon ausgehen, dass der Test aufgrund dieser Vermutung durchgeführt wird. Konsequenterweise ist also der Test aus Szenario 2 durchzuführen. Bezeichnenderweise haben auch die Betreiber der URL https://www.mathe-aufgaben.com ihren (im Sinne der Aufgabensteller(inne)n formulierten) Lösungshinweisen die Bemerkung hinzugefügt, dass die ersten beiden Sätze der Aufgabenstellung aus Sicht der Autoren leider verwirrend gestellt seien.

Diese verwirrende Formulierung, die von verschiedenen Lehrkräften bemerkt und mir zugetragen wurde, ist aber wohl von den Aufgabensteller(inne)n bewusst gewählt worden, da beim Übergang vom Kernfach Mathematik zum Leistungsfach ein automatisierter Testalgorithmus zugunsten eines genaueren Hinsehens sowie Argumentierens und Hinterfragens aufgebrochen werden sollte. Dieser Versuch endete jedoch in einem – wie es eine Lehrkraft formuliert hat – „logisch-stochastischen Chaos".

Das gleiche Chaos wurde mit Aufgabe C 2.2 angerichtet. Auch hier wird die Vermutung als Nullhypothese formuliert, und sie kann deshalb prinzipiell nicht statistisch abgesichert werden.

3 Fazit

Es ist seit langem bekannt, dass selbst *Studierende* der empirischen Wissenschaften meist nicht verstanden haben, was man aus einem Testergebnis überhaupt schließen kann. Wenn ein Test zum 1%-Niveau einer Nullhypothese H_0 gegen eine Alternativhypothese H_1 die Nullhypothese abgelehnt hat, so ist jede der folgenden Antwortmöglichkeiten eines Studierenden der Psychologie (leicht modifiziert) vorgelegten Fragebogens (s. Krauss und Wassner (2001)) falsch, aber jede(r) Befragte hatte mindestens eine Antwort als richtig angekreuzt.

1) Es ist eindeutig bewiesen, dass H_0 falsch ist.
2) Die Wahrscheinlichkeit des Zutreffens von H_0 ist gefunden worden.
3) Es ist eindeutig bewiesen, dass H_1 wahr ist.
4) Man kann nun die Wahrscheinlichkeit ableiten, dass H_1 richtig ist.
5) Entscheidet man sich nun, H_0 zu verwerfen, dann kennt man jetzt die Wahrscheinlichkeit dafür, dass diese Entscheidung falsch ist.
6) Wenn man das Experiment sehr oft wiederholen würde, würde man in 99% der Fälle ein signifikantes Ergebnis bekommen.

Für die Schule ähnlich niederschmetternde Ergebnisse finden sich in Mossburger (2014). Es ist also eine Illusion zu glauben, Schülerinnen und Schüler der gymnasialen Kursstufe würden *verstehen*, was ein Signifikanztest überhaupt leisten kann und was nicht.

Das statistische Testen – zumindest so, wie es noch in vielen Bundesländern im Gymnasium behandelt wird – läuft weitgehend rezeptartig ab und hat nur wenig mit Mathematik zu tun. Überdies sind die in der Schule verwendeten Einkleidungen meist an den Haaren herbeigezogen (s. Henze (2020a)) und tragen somit zumindest prinzipiell zur Diskreditierung der Mathematik bei. Einige Bundesländer haben bereits statistische Tests aus den Lehrplänen genommen und stattdessen für den Bereich der schließenden Statistik Konfidenzintervalle eingeführt. Ob Letztere jedoch von Schülerinnen und Schülern und den für das Stellen von Aufgaben Verantwortlichen wirklich *verstanden* werden, bleibt abzuwarten.

Was die beiden diskutierten Abituraufgaben betrifft, so dürften diese – nach dem Zitat einer Lehrkraft – „den Kolleg*innen auch noch die letzte Motivation für das ohnehin schon unbeliebte Thema der Hypothesentests nehmen". Wenn der Unterschied vom Kerncurriculum zum Leistungsfach darin besteht, dass man darüber spekulieren muss, wer nun einen Test in Auftrag gibt, und wenn man erfährt, dass Vermutungen als Nullhypothesen zu formulieren sind, obwohl der gesunde Menschenverstand das Gegenteil fordert, muss man sich fragen, ob es – wiederum Zitat einer Lehrkraft – „um Sprachklauberei geht, die nichts mehr mit Mathematik zu tun hat und alle nur vergrault". Dabei gibt es zahlreiche spannende Themen für einen verständnisorientierten Stochastikunterricht. Einige davon finden sich in Henze u.a. (2021). Möchte man etwa abseits vom schematischen Testen von Schein-Hypothesen Schülerinnen und Schülern auf eindrucksvolle Weise demonstrieren, dass die Stochastik auch zum Klimawandel einiges zu sagen hat, kann man etwa Rekorde bei Temperaturdaten analysieren (siehe z.B. Kap. 5 in Henze u.a. (2021) und Henze (2020b) sowie Henze (2020c)).

Wir sollten uns fragen: Was würde man verlieren und was gewinnen, wenn man die statistischen Tests aus den gymnasialen Lehrplänen streicht und deren Vermittlung den Universitäten überlässt?

Danksagung: Ich danke den beiden Gutachtern sowie Joachim Engel für wertvolle Hinweise.

Literatur

Dürr, R., u.a. (2017): Lambacher Schweizer. Mathematik für Gymnasien Baden–Württemberg. Kursstufe. Stuttgart u.a.: Klett.

Henze, N. (2020a). Über Sinn und Unsinn statistischen Testens. Erklärvideo. https://publikationen.bibliothek.kit.edu/1000121006

Henze, N. (2020b). Rekorde in zufälligen Permutationen - Teil 1. Erklärvideo. DOI:10.5445/DIVA/2020-108

Henze, N. (2020c). Rekorde in zufälligen Permutationen - Teil 2. Erklärvideo. DOI:10.5445/IR/1000118548

Henze, N. (2021): Stochastik für Einsteiger. 13. Auflage. Heidelberg: Springer Spektrum.

Henze, N., Müller, K., und Schilling, J. (2021). Stochastik rezeptfrei unterrichten – Anregungen für spannende Lehre über den Zufall: Heidelberg: Springer Spektrum.

Krauss, St., und Wassner, Ch. (2001). Wie man das Testen von Hypothesen einführen sollte. Stochastik in der Schule 21(1), S. 29–34.

Mossburger, M. (2014). Unklare Begriffe und Wunschdenken bei Signifikanztests. Stochastik in der Schule 34(1), S. 2–8.

Anschrift des Verfassers:

Prof. i.R. Dr. Norbert Henze
KIT Distinguished Senior Fellow
Institut für Stochastik
Karlsruher Institut für Technologie (KIT)
Englerstr. 2
76131 Karlsruhe
`Henze@kit.edu`