

# Auf der Suche nach $H_0$

## Signifikanztests und das Risiko falscher Entscheidungen

**LERNGRUPPE:** 12. Schuljahr

**IDEE:** Mit einem Entscheidungsspiel die „Philosophie“ einseitiger Signifikanztests entdecken

**ARBEITSBLATT:** Spielstrategien analysieren

**WEITERES MATERIAL:** Beutel mit roten/weißen Kugeln

**VORWISSEN:** Pfadregel, Binomialverteilung, evtl. Regel von Bayes

**ZEITBEDARF:** 4 Stunden

In vielen Bundesländern wird „beurteilende Statistik“ durch das einseitige Testen von Hypothesen abgedeckt. Die Wahl der Nullhypothese  $H_0$  und der Alternative  $H_1$  bringt nicht nur Lernende oft zur Verzweiflung, weil sie von (meist versteckt formulierten) Interessenlagen abhängt, die bei der Abiturvorbereitung aus eingekleideten Aufgabentexten zu decodieren sind. Wenn man aber die Interessen durch explizite Bewertung von Fehlentscheidungen im Rahmen eines authentischen Gewinnspiels offenlegt und berechenbar macht, werden die unterschiedlichen Rollen der Hypothesen transparent. Schülerinnen und Schüler entdecken so die hinter Rezepten versteckten Konzepte.

Eine Spielsituation hat dabei einen lernpsychologischen Vorteil: Man kann den Hypothesen A-priori-Wahrscheinlichkeiten zuordnen und in Abhängigkeit von Testergebnissen A-posteriori-Wahrscheinlichkeiten berechnen. So wird quantifizierbar, was es bedeutet, Hypothesen auf dem 5%-Signifikanzniveau (mit 5% Irrtumswahrscheinlichkeit) zu verwerfen: Die verworfene Hypothese  $H_0$  wird unwahrscheinlicher, aber die Alternative  $H_1$  gilt weit verbreiteten Fehlvorstellungen zum Trotz

beileibe nicht mit 95%iger Sicherheit. Last but not least führt das Verlassen der Spielsituation zu einer kritischen Bewertung der häufig völlig überschätzten Aussagekraft von Signifikanztests und wirkt daher im besten Sinne allgemeinbildend.

### Einseitige Hypothesentests

Zur Einstimmung auf das Thema erinnert ein Beispiel (Alkohol-Stichprobe, Bamberg 2009) an das Ritual, nach dem Hypothesentests ablaufen:

Eine Fabrik erzeugt Alkopop-Getränke mit einem Alkoholgehalt von 5%. Dabei treten beim Alkoholgehalt Schwankungen auf. Die Hypothese  $H_0$ , der Alkoholgehalt sei gleich dem Sollwert  $p_0 = 0,05$ , soll anhand einer Stichprobe überprüft werden. Aufgrund der *Interessenlage* der Personen, die die Untersuchung vornehmen, sind drei Fälle zu unterscheiden: Die Überprüfung geschieht durch

a) eine Eichkommission, die an einer Abweichung vom Sollwert  $p_0 = 0,05$  sowohl nach unten als auch nach oben interessiert ist,

b) eine Verbraucherorganisation, die daran *interessiert* ist, dass tatsächlich die 5% Alkohol in den Flaschen sind. Sie stellt *misstrauisch* die Frage, ob der wahre Alkoholgehalt *kleiner* als der Sollwert ist,

c) eine Jugendschutzkommission, die *befürchtet*, dass zu viel Alkohol in den Flaschen ist, um die Konsumenten möglichst schnell alkoholabhängig zu machen. Sie stellt – ebenfalls *misstrauisch* – die Frage, ob der wahre Alkoholgehalt *größer* als der Sollwert ist.

Der Alkoholinhalt 5% ist zwar keine Wahrscheinlichkeit, aber der Kontext ist authentischer als die Untersuchung von Hypothesen, die aus Gründen

geforderter Aufgabenformate meist „aus dem Hut gezaubert“ werden (wie etwa die berühmte „Reinlich & Sohn“-Aufgabe aus dem NRW-Abitur 2007, Stoyan 2011). Diesen Kontext kann man wie folgt in Testverfahren übersetzen (Bamberg 2009):

$$H_0: p = p_0, \quad H_1: p \neq p_0$$

$$H_0: p \leq p_0, \quad H_1: p > p_0$$

$$H_0: p \geq p_0, \quad H_1: p < p_0$$

Dabei ist aber aus der Aufgabenstellung gar nicht klar ersichtlich, mit welchem Ziel die Beteiligten ihre Untersuchungen anstellen. So könnten die Jugendschützer, die beruhigt schlafen wollen, versuchen,  $H_0: p = 0,05$  zugunsten von  $H_1: p < 0,05$  zu verwerfen, andere, die erfolgreich klagen wollen,  $H_0: p = 0,05$  zugunsten von  $H_1: p > 0,05$  zu verwerfen. Mit welcher Sorte Jugendschützer man es zu tun hat, ist einzig aus dem Signalwort „*misstrauisch*“ zu erraten. Gleiches gilt für die Verbraucherschützer.

Der gesunde Menschenverstand fragt sich aber: Ist nicht der zweiseitige Test der Eichkommission unabhängig von unterstellten Interessenlagen für Verbraucher- und Jugendschützer gleichermaßen interessant? Schließlich läuft ein zweiseitiger Test auf dem 10%-Signifikanzniveau rechnerisch auf einen einseitigen auf dem 5%-Niveau hinaus.

Warum sollte man also vorher über die Alternativen spekulieren? Wenn der Alkoholgehalt unterhalb des zweiseitigen Prognoseintervalls liegt, werden die Verbraucherschützer klagen und die Jugendschützer beruhigt schlafen; wenn er oberhalb liegt, werden die Verbraucherschützer ruhig schlafen und die Jugendschützer klagen. Wenn der Alkoholgehalt innerhalb des 90%-Prognoseintervalls liegt, wird man die Sache auf sich beruhen lassen – oder größere

Stichproben nutzen, bis evtl. irrelevant kleine Abweichungen signifikant werden. (Trotz der allgegenwärtigen Forderung nach Allgemeinbildung in Lehrplan-Präambeln wird der Unterschied zwischen Signifikanz und Relevanz derzeit in keinem Bundesland thematisiert).

Aus didaktischer Sicht – da sind sich im Gegensatz zu Lehrplankommissionen viele Fachleute einig – spricht vieles dafür, in der Schule auf einseitige Hypothesentests zu verzichten (Buth u. a. 2002). Sinnvoller ist die *Arbeit mit Konfidenzintervallen*, weil man (im Gegensatz zu Testverfahren) den Einfluss des Stichprobenumfangs an der Länge des Konfidenzintervalls unmittelbar ablesen kann – und weil Konfidenzintervalle die Idee des Messens und der Messungenauigkeit aus der Geometrie (Messen von Längen) in die Stochastik (Messen von Wahrscheinlichkeiten) übertragen (vgl. hierzu Henze u. a. 2020).

## Konzeptvergleich

In Bezug auf die Wirklichkeit haben alle Wahrscheinlichkeitsangaben hypothetischen Charakter. Sie sind stets nur Modelle, mit denen man versucht, die Wirklichkeit zu beschreiben, die aber „nie“ genau stimmen. *Zweiseitige Tests* und *Konfidenzintervalle* sind Werkzeuge, mit deren Hilfe man (vorurteilsfrei, unter Einsatz von Fingerspitzengefühl und Sachverstand) die Güte von Modellen vergleichen und bewerten kann. Sie helfen bei der Suche nach besseren Modellen.

Hinter *einseitigen* Signifikanztests steckt ein anderes Grundkonzept. Es geht nicht mehr um die Suche nach Wahrheit, sondern um das Entscheiden zwischen verschiedenen Modellen mit dem Ziel einer Optimierung des Nutzens oder der Begrenzung des Schadens.

Diesen prinzipiellen Unterschied unter dem Motto „Gewinn schlägt Wahrheit“ am eigenen Leibe (im Unterricht tatsächlich mit Glücksgefühlen bei richtigen und Frusterlebnissen bei falschen Tipps) zu erleben, ist das Anliegen des Entscheidungsspiels aus der Kopiervorlage. Im Rahmen des Spiels lassen sich Bewertungen von Fehlentscheidungen gezielt variieren und die zugehörigen

optimalen Entscheidungsregeln bestimmen.

Man erkennt, dass die Entscheidungsregeln mit der größten Gewinnerwartung nicht die meisten richtigen Entscheidungen garantieren. Aber sie halten die Wahrscheinlichkeit gravierender Fehler klein. Man entdeckt und erlebt beim einseitigen Test die Bedeutung von Nullhypothese und  $\alpha$ -Fehler.

Ein Transfer der Erkenntnisse auf Realsituationen liegt nahe, und es wird deutlich, auf welche (von den meisten Anwendern herbeigewünschten) Aussagen man bedauerlicherweise verzichten muss, wenn man beim einseitigen Hypothesentest wichtige Parameter nicht kennt und – wie in einem Blindflug im Nebel – trotzdem gezwungen wird, Entscheidungen zu fällen.

## Ein Entscheidungsspiel

### Entscheidungsspiel

Man spielt in Tandems (**Arbeitsblatt 1**). Der erste Partner, „Hannah“, wählt zufällig eine Trefferwahrscheinlichkeit A (mit  $p < \frac{1}{2}$ ) oder B ( $p > \frac{1}{2}$ ) und nennt die Ergebnisse einer Bernoulli-Kette fester Länge  $n$ . Der zweite Partner „Florian“ muss auf A oder auf B tippen. Der gesunde Menschenverstand tippt auf A, wenn die relative Trefferhäufigkeit  $h$  unter  $\frac{1}{2}$  liegt, sonst auf B. Wenn die beiden Alternativen A und B spezifiziert sind und (wegen der Zufallsauswahl) selber mit bekannter A-priori-Wahrscheinlichkeit auftreten (in der Spielsituation treten A:  $p = \frac{1}{3}$  und B:  $p = \frac{2}{3}$  je mit Wahrscheinlichkeit  $\frac{1}{2}$  auf, s. **Abb. 1**), ist die Situation mit der Binomialverteilung komplett berechenbar. Bei  $n = 11$  beträgt die Wahrscheinlichkeit richtiger Tipps 88 %, und die Gegenwahrscheinlichkeit 12 % ist tatsächlich (im eigentlichen Wortsinn) die Irrtumswahrscheinlichkeit, also die Wahrscheinlichkeit dafür, dass man sich bei der Entscheidung irrt.

### Gewinnspiel

Durch Belohnung richtiger und Bestrafung falscher Entscheidungen wird aus dem Entscheidungsspiel ein Gewinnspiel, dessen Gewinnerwartung man gemäß **Arbeitsblatt 1** leicht berechnen kann.

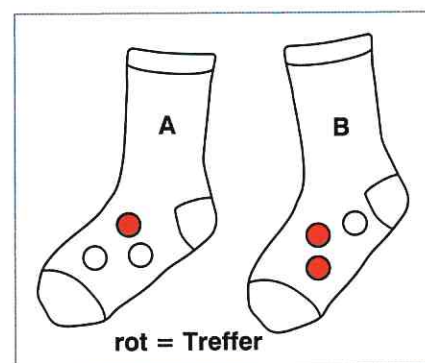
Wenn man nach einer symmetrischen Bestrafung von Fehlentscheidungen zu einer unsymmetrischen übergeht, dann hat man den Schlüssel für ein selbstständiges Entdecken und Analysieren der Grundidee einseitiger Hypothesentests. So werden in der Kopiervorlage richtige Tipps stets mit 1P belohnt, der falsche Tipp auf B stets mit  $-2P$  ... und der falsche Tipp auf A nacheinander mit  $-2P$ ,  $-20P$ ,  $-40P$ ,  $-100P$  ... bestraft.

Der gesunde Menschenverstand legt dann nahe, im Zweifel eher auf B zu tippen als auf A. In der Sprache der Signifikanztests wächst der Annahmehereich von B:

B wird zur Nullhypothese  $H_0$ , bei der man „konservativ“ möglichst lange bleiben sollte und deren fälschliches Verwerfen man vermeiden sollte.

Dass diese Strategie aufgeht, zeigen Experimente und Simulation: Bei symmetrischer Bestrafung von Fehlentscheidungen (je  $-2P$ ) und der Strategie „Tippe auf A:  $p = \frac{1}{3}$  bei  $X \leq 5$  (von 11) Treffern, sonst auf B“ wächst der Gewinn bei einer Folge von Spielen stetig, mit kleinen Rückschlägen. So hat man in der Simulation aus **Abb. 2** in 100 Spielen 86-mal richtig getippt, 5-mal falsch auf A, 9-mal falsch auf B und dabei  $86 \cdot 1 + 5 \cdot (-2) + 9 \cdot (-2) = 58$  Punkte gewonnen. Bei  $b = -20$  hätte man bei der gleichen Tippstrategie und gleichem Spielverlauf  $86 \cdot 1 + 5 \cdot (-2) + 9 \cdot (-20) = -104$  Punkte verloren (**Abb. 3**).

Die vorsichtigeren Strategie „Tippe auf A:  $p = \frac{1}{3}$  bei  $X \leq 4$  (von 11) Treffern“



**Abb. 1:** Im Spiel werden die Trefferwahrscheinlichkeiten im Urnen-Modell realisiert.

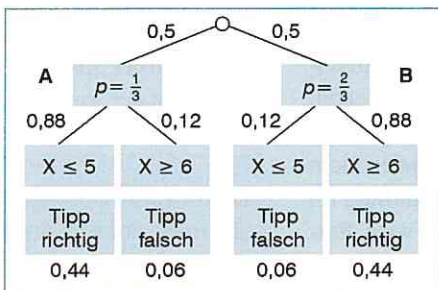
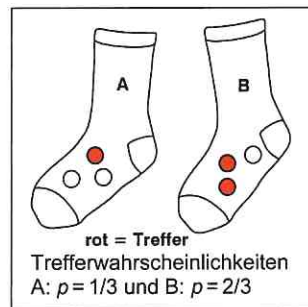


# Das Risiko begrenzen – Gewinn schlägt Wahrheit

## 1. Entscheidungsspiel

Hannah zieht zufällig eine der Socken A oder B. Dann zieht sie daraus  $n = 11$ -mal mit Zurücklegen und nennt die Ergebnisse (rote Kugel = Treffer), zum Beispiel T N N T N N T T T N (das heißt, 5 von 11 Ergebnissen sind Treffer).

Florian muss tippen, ob Hannah Socke A oder Socke B erwischt hat. Er muss sich also bzgl. der Trefferwahrscheinlichkeit zwischen den Hypothesen A:  $p = 1/3$  und B:  $p = 2/3$  entscheiden.



### Regel von Bayes

Nachdem man B (wegen  $X \leq 5$ ) auf dem Signifikanzniveau  $\alpha = 12\%$  verworfen hat, gilt A mit der Wahrscheinlichkeit 0,44:  $(0,44 + 0,06) = 88\%$ .

für die optionale Aufgabe d)

- Der gesunde Menschenverstand rät Florian, bei  $X \leq 5$  Treffern auf Socke A:  $p = 1/3$  zu tippen und bei  $X \geq 6$  Treffern auf Socke B. Erproben Sie diese Strategie (wir nennen sie A5) und schätzen Sie (in %), wie häufig Florian dann richtig tippt.
- Berechnen Sie die Wahrscheinlichkeit dafür, dass diese Strategie A5 richtige Tipps liefert.
- Man könnte auch bei  $X \leq 4$  Treffern auf A und bei  $X \geq 5$  Treffern auf B oder aber bei  $X \leq 6$  Treffern auf A und bei  $X \geq 7$  Treffern auf B tippen. Zeigen Sie, dass diese Strategien (wir nennen sie A4 und A6) auf lange Sicht weniger richtige Tipps liefern.
- Exkurs (Bayes): Da Hannah zufällig auswählt, ordnet Florian vor dem Versuch der Socke A die subjektive A-priori-Wahrscheinlichkeit 50 % zu. Berechnen Sie mithilfe der Bayes'schen Regel, auf welchen Wert Florians Wahrscheinlichkeit für A ansteigt, wenn Hannah eine Trefferzahl  $X \leq 5$  meldet. (Man bezeichnet diese als A-posteriori-Wahrscheinlichkeit).

## 2. Gewinnspiel

### a) Symmetrische Bestrafung von Fehlentscheidungen

Aus dem Entscheidungsspiel wird ein Gewinnspiel, wenn man – zum Beispiel – bei jedem richtigen Tipp 1 P gewinnt und für jeden falschen Tipp 2 P verliert. Suchen Sie die Tippstrategie, die dann maximalen Gewinn sichert. Bestimmen Sie den maximalen Gewinn.

### b) Asymmetrische Bestrafung von Fehlentscheidungen

Suchen Sie die Strategie, die den maximalen Gewinn (minimalen Verlust) garantiert, wenn

- wie in Aufgabenteil a) ein richtiger Tipp 1 P einbringt, aber
- ein falscher Tipp auf B (also das fälschliche Verwerfen von A) 2 P kostet ( $a = -2$ ) und
- ein falscher Tipp auf A (also das fälschliche Verwerfen von B) 20 P bzw. 40 P bzw. 60 P kostet ( $b = -20$  bzw.  $b = -40$  bzw.  $b = -60$ ).

Bestimmen Sie jeweils den maximalen Gewinn.

- Fassen Sie die Ergebnisse tabellarisch zusammen und erläutern Sie die Bedeutung der Überschrift: „Das Risiko begrenzen – Gewinn schlägt Wahrheit“.

a	b	optimale Strategie	erwarteter Gewinn	richtige Tipps	B wird fälschlich verworfen mit	A wird fälschlich verworfen mit	nach Verwerfen von B gilt a-posteriori für A
-2	-2						
-2	-20						
-2	-40						
-2	-69						

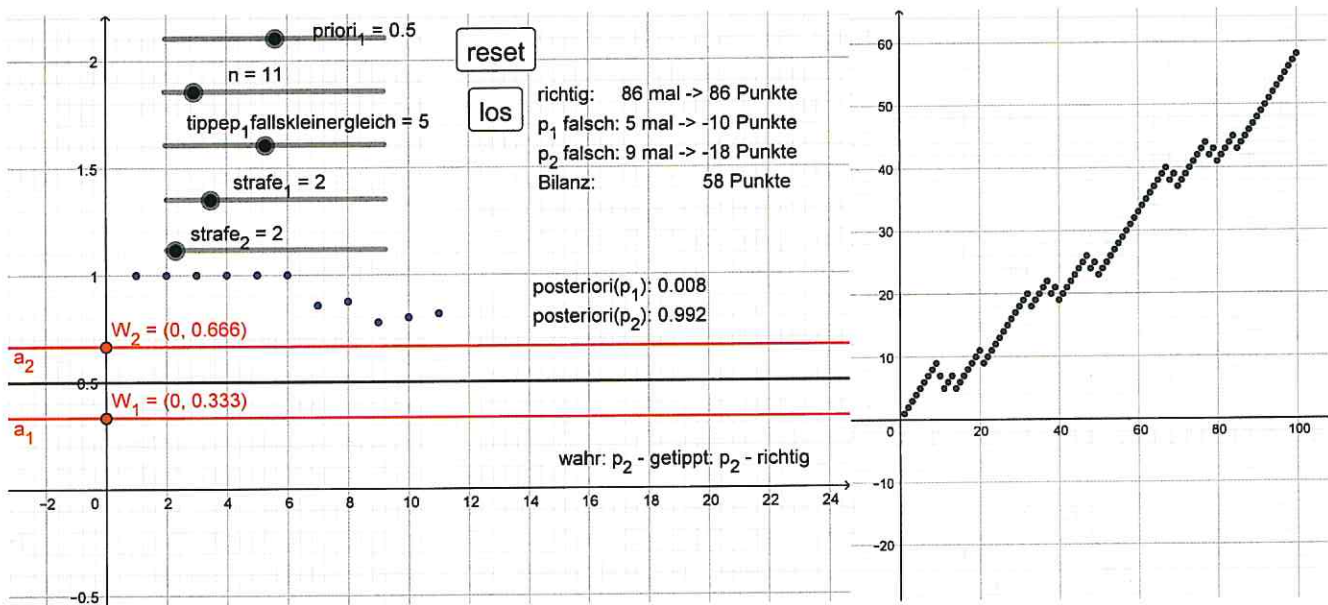


Abb. 2: Gewinnverlauf bei symmetrischer Bestrafung

tippe A bei	richtige Tipps	Gewinnerwartung bei priori (A) = 0.5					$\alpha$	P(A)	$\alpha + P(A)$
		$b = -2$	$b = -20$	$b = -40$	$b = -60$	$b = -100$			
$X \leq 2$	61.6%	-0.151	-0.163	-0.177	-0.191	-0.218	0.1%	99.4%	99.6%
$X \leq 3$	73.2%	0.196	0.116	0.028	-0.060	-0.237	0.9%	98.2%	99.0%
$X \leq 4$	83.6%	0.509	0.161	-0.225	-0.612	-0.384	3.9%	94.8%	98.7%
$X \leq 5$	87.8%	0.634	-0.465	-1.686	-2.907	-5.384	12.2%	87.8%	100.0%
$X \leq 6$	83.6%	0.509	-2.092	-4.982	-7.872	-13.652	28.9%	76.9%	105.8%

Tab. 1.: Markiert ist die optimale Gewinnerwartung bei A-priori-Gleichverteilung.

tippe A bei	richtige Tipps	Gewinnerwartung bei priori (A) = 0.8					$\alpha$	P(A)	$\alpha + P(A)$
		$b = -2$	$b = -20$	$b = -40$	$b = -60$	$b = -100$			
$X \leq 2$	38.7%	-0.839	-0.844	-0.849	-0.855	-0.866	0.1%	99.9%	100.0%
$X \leq 3$	57.6%	-0.271	-0.303	-0.338	-0.374	-0.444	0.9%	99.5%	100.4%
$X \leq 4$	76.1%	0.283	0.144	-0.010	-0.165	-0.474	3.9%	98.7%	102.5%
$X \leq 5$	87.8%	0.634	0.194	-0.294	-0.782	-1.759	12.2%	96.6%	108.8%
$X \leq 6$	91.1%	0.734	-0.306	-1.462	-2.618	-4.930	28.9%	93.0%	121.9%

Tab. 2.: Markiert ist die optimale Gewinnerwartung bei priori (A) = 0.8.

screenshot: GeoGebra/GeoGebra.org

hätte bei gleichem Spielverlauf immerhin noch 81 Treffer und einen Gewinn von +25 Punkten erbracht (Abb. 4). Mit Hilfe des Baumdiagramms aus dem Arbeitsblatt und der Binomialverteilung berechnet man für verschiedene Tippstrategien die Wahrscheinlichkeiten

richtiger Tipps (Tab. 1, Spalte 2) und die Gewinnerwartungen bei verschiedenen Bestrafungen  $b$  der falschen Tipps auf A – das heißt des fälschlichen Verwerfens von B – (Tab. 1, Spalten 3 bis 7). Die für jede der genannten Bestrafungen maximal möglichen Gewinnerwartungen

sind in den Spalten 3 bis 7 farbig unterlegt. Spalte 8 zeigt, auf welchem Signifikanzniveau  $\alpha$  man beim Tipp auf A die Hypothese  $H_0$  verwirft. Spalte 9 zeigt die Wahrscheinlichkeit, mit der (bei vorausgesetzter A-priori-Gleichwahrscheinlichkeit beider Hypothesen) nach



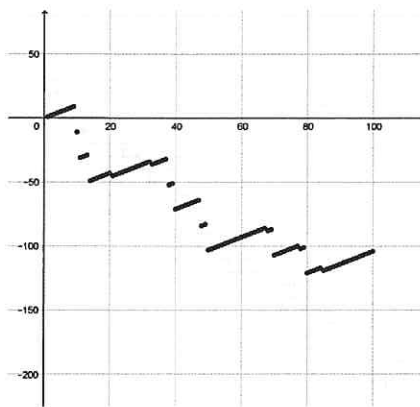


Abb. 3: Fehllentscheidungen werden bestraft

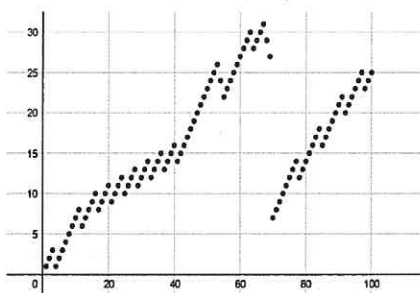


Abb. 4: Die vorsichtigeren Strategie gewinnt

Verwerfen von  $H_0$  dann  $H_1$  gilt. Spalte 10 dokumentiert, dass A ( $H_1$ ) nach dem Verwerfen von B ( $H_0$ ) nicht mit der Gegenwahrscheinlichkeit des Signifikanzniveaus  $(1 - \alpha)$  gilt.

Man erkennt: Je stärker man das fälschliche Verwerfen von B ( $H_0$ ) durch  $b = -2, -20, -40, -60, -100$  bestraft, desto konservativer sollte man sich bei Tipps auf A ( $H_1$ ) verhalten. Bei  $b = -40$  ist mit Tipps auf A im Falle  $X \leq 3$  langfristig noch ein kleiner Gewinn zu erwirtschaften. Bei noch höheren Bestrafungen kann man aber auch bei noch vorsichtigeren Tippstrategien langfristig nur verlieren.

Wenn A ( $H_1$ ) a-priori wahrscheinlicher wird (80 % statt 50 %, indem man in der Kopiervorlage vier Socken der Sorte A und nur eine der Sorte B zur Auswahl anbietet), braucht man beim Tipp auf A nicht mehr ganz so vorsichtig sein, um den Gewinn zu maximieren. Die Veränderungen dokumentiert **Tab. 2**.

## Aus Spiel wird Ernst

Das Spiel funktioniert natürlich auch, wenn der erste Partner die Trefferwahrscheinlichkeit  $p$  für seine Bernoulli-

Kette „frei Schnauze“ bzw. durch einen unbekanntem Zufallsgenerator festlegt (also nicht mehr nur mit  $p = 0,5$  zwischen  $\frac{1}{3}$  und  $\frac{2}{3}$  wählt) und der zweite Partner unter Androhung von Strafe bei falschen Entscheidungen gezwungen wird, nach Bekanntgabe der Trefferzahl  $X$  auf A:  $p < 0,5$  oder B:  $p > 0,5$  zu tippen.

Genau das ist die Situation, in der sich Schülerinnen und Schüler wiederfinden, wenn sie Abituraufgaben zu einseitigen Signifikanztests zu bearbeiten haben. Sie müssen aus dem Aufgabentext herauschälen, ob der Aufgabenautor das fälschliche Verwerfen von A oder das fälschliche Verwerfen von B für „schlimmer“ hält und stärker „bestraft“. Dabei sind

- weder Informationen über mögliche Trefferwahrscheinlichkeits-Alternativen,
- schon gar nicht über deren A-priori-Gültigkeit,
- und auch keine Informationen über die Stärke der Bestrafung von Fehlentscheidungen verfügbar.

Max, der im GK 12 die Zusammenhänge anhand des Entscheidungsspiels durchschaut hatte, konstatierte lakonisch, dass „man dann eigentlich nichts Interessierendes mehr berechnen kann“, und brachte damit die Situation auf den Punkt. Noch kritischer formuliert: „Signifikanztests werden durchgeführt, wenn man fast keine Information über die vorliegende Situation hat, mit der Folge, dass man sich mit Antworten auf falsch gestellte Fragen zufriedengeben muss, aber auf die eigentlich interessierenden Fragen (nach der Wahrscheinlichkeit, mit der Hypothesen gelten) keine Antwort erhält“ (Buth 2002).

Tatsächlich ist man bei einseitigen Signifikanztests gezwungen, der im Entscheidungsspiel als tragfähig entdeckten Strategie in ritualisierter Weise weiterhin zu folgen: Falls man das fälschliche Verwerfen von  $H_0$  für „schlimm“ hält, testet man auf dem 5 %-Signifikanzniveau, und falls man es für „sehr schlimm“ hält, auf dem 1 %-Signifikanzniveau. Falls man dann die Hypothese verwerfen kann, weiß man aber rein gar nichts über die Wahrscheinlichkeit, mit der sie (oder eine der Alternativen) gilt. Eine Hypothese zu verwerfen bedeutet de facto nichts anderes, als dass man ein „schlechtes

Gewissen“ oder ein „mulmiges Bauchgefühl“ haben sollte, wenn man trotzdem bei ihr bleibt, ... und lieber weitere Untersuchungen anstellen sollte, als routinemäßig auf „die Alternative“ zu vertrauen.

## Resümee

Das Entscheidungsspiel ermöglicht durch transparente Bewertung von Fehlentscheidungen das selbstständige (daher lernpsychologisch nachhaltige) Entdecken des Konzepts, das hinter einseitigen Signifikanztests steckt – und der unterschiedlichen Rollen, welche die Nullhypothese  $H_0$  und deren Alternative  $H_1$  dabei spielen.

Über die im Spielkontext naheliegende Bewertung der Gültigkeit von Hypothesen durch A-priori-Wahrscheinlichkeiten gelingt es, naheliegende und weitverbreitete Fehlvorstellungen bei der Interpretation signifikanter Testergebnisse aufzuspüren: Bei nicht spezifizierter Alternative und fehlender A-priori-Bewertung weiß man NICHTS über die Gültigkeit einer Hypothese, wenn ein Ergebnis im Ablehnungsbereich liegt. „Mathematikunterricht sollte aufklären, wie wenig ein Signifikanztest aussagt“ (Mossburger 2013).

Zu dieser Aufklärung leistet das Entscheidungsspiel einen substanziellen Beitrag.

## Hinweis

Die verwendeten Programme sind im Download zu diesem Heft verfügbar. Gleiches gilt für eine ausführliche Kopiervorlage, die das Entscheidungsspiel elementarer mit (händischen) Simulationen, ohne bedingte Wahrscheinlichkeiten (Regel von Bayes) und ohne Bezug zu Signifikanztests nutzt, um Risiken beim Entscheiden unter Unsicherheit schon in der Sek I erleben und erforschen zu lassen.

## Literatur

- Bamberg, G./Baur, F./Krapp, M. (2009): Statistik. München.
- Buth, M. (2002): Anmerkungen zum Testen von Hypothesen – In: Stochastik in der Schule 2002/2, S. 27–29.
- Henze, N./Hotz, T./Riemer, W./Skorsetz, B./Vehling, R. (2020): Schickt die statistische Signifikanz in den Ruhestand. – In: Der Mathematikunterricht 4/2020.
- Mossburger, M. (2014): Unklare Begriffe und Wunschdenken bei Signifikanztests. – In: Stochastik in der Schule 2014/1, S. 2–8.
- Stoyan, D. (2011): Statistische Tests in Gymnasiallehrbüchern. – In: Stochastik in der Schule 2011/1, S. 28–32.