

Das Formulieren der Nullhypothese beim Signifikanztest mit Blick auf eine authentische Anwendung

ANNA SCHÄFER, PADERBORN

Zusammenfassung: Welche Schwierigkeiten beim Formulieren der Nullhypothese auftreten und wodurch diese bedingt sein könnten, wurde exemplarisch anhand von Studierendenbearbeitungen zu einer authentischen Anwendung eines Signifikanztests untersucht. Darauf aufbauend wird ein Vorschlag unterbreitet, wie man diesen Schwierigkeiten durch die Reflexion des Modellierungsprozesses, die bewusste Unterscheidung zwischen der substantiellen Hypothese und der statistischen Nullhypothese sowie die Erarbeitung von Kriterien für gute Nullhypothesen entgegenwirken kann.

1 Einleitung

In einer Zeit, in der uns statistische Informationen täglich in Form von Grafiken, Prozentzahlen und signifikanten Studienergebnissen präsentiert werden, ist „die Fähigkeit, statistische Informationen, datenbasierte Argumente und stochastische Phänomene in unterschiedlichen Kontexten zu interpretieren und kritisch zu bewerten“ (Krüger 2016, S. 3) von enormer Bedeutung. Um bei Lernenden eine angemessene „statistische Grundbildung“ (ebd., S. 2 f) zu erreichen, muss nicht nur statistisches Grundwissen gefördert werden, sondern auch die Fähigkeit, dieses in geeigneten Sachsituationen anzuwenden. Dazu bietet es sich an, mit realen Daten zu arbeiten (vgl. Engel 2007). Wohingegen es im Bereich der Beschreibenden Statistik schon zahlreiche Unterrichtsvorschläge gibt, ist es schwieriger, authentische Beispiele¹ zu finden, die das Testen von Hypothesen auf Schulniveau behandeln. Ein solches wird unter Abschnitt 3 vorgestellt. Die Behandlung von authentischen Anwendungen kann bei den Lernenden ferner Fehlvorstellungen und Missverständnisse bezüglich der Grundidee des Testens von Hypothesen aufdecken, die bei eingekleideten Aufgaben vermutlich nicht aufgefallen wären. Dass es diesbezüglich viele Schwierigkeiten gibt, wurde in der Literatur bereits breit diskutiert und empirisch belegt (vgl. z. B. Gigerenzer & Krauss 2001; Nickerson 2000). Dabei wird in der Regel Testergebnissen mehr Aussagekraft zugesprochen, als diese eigentlich haben. Beispielsweise wird der P-Wert als Wahrscheinlichkeit für die Gültigkeit der Nullhypothese und die Gegenwahrscheinlichkeit zum P-Wert als Wahrscheinlichkeit angesehen, dass die Ergebnisse replizierbar sind. Diese falschen Vorstellungen treten nicht nur bei Lernenden an Schulen und Universitäten auf, sondern

sind auch unter Wissenschaftlern verbreitet. In diesem Artikel² soll eine weitere Fehlvorstellung im Fokus stehen, nämlich, dass man aus dem Verwerfen der statistischen Nullhypothese darauf schließen kann, dass die außermathematische Theorie, nach der die Nullhypothese falsch sein müsste, richtig ist. Dieser Schluss ist jedoch allein aus dem statistischen Testergebnis nicht möglich (vgl. Nickerson 2000, S. 254).

Im Folgenden soll zunächst der Anwendungskontext der „verlorenen Mädchen von Gorleben“ sowie die Lehrveranstaltung, in dessen Rahmen die Aufgabe bearbeitet wurde, vorgestellt werden.

2 Einordnung in die Lehrveranstaltung

Die Lehrveranstaltung „Elemente der Stochastik“³ an der Universität Paderborn im Sommersemester 2015 wurde von ca. 100 Bachelor-Studierenden des Lehramts für Haupt- und Real- und Gesamtschulen (HRGe) besucht. Laut Studienverlaufsplan soll die Veranstaltung im 4. Fachsemester besucht werden. Zuvor sollten die Studierenden bereits die Elemente der Arithmetik, der Geometrie und der Analysis sowie z. T. zugehörige Didaktikvorlesungen erfolgreich abgeschlossen haben. Da es jedoch keine formalen Teilnahmevoraussetzungen gibt, nahmen letztendlich Studierende vom 1. bis zum 8. Fachsemester teil. Zur Stochastik bringen die Studierenden tendenziell wenige Vorerfahrungen aus der Schule mit. Circa die Hälfte der Studierenden gab in einer schriftlichen Befragung an, dass sie in der Oberstufe gar keinen Stochastikunterricht hatten bzw. sich nicht daran erinnern könnten.

In der Veranstaltung wurden drei Themengebiete der Stochastik bearbeitet: Vor einem Einblick in die Beurteilende Statistik wurden die Beschreibende Statistik sowie Grundlagen der Wahrscheinlichkeitsrechnung ausführlich behandelt. Insbesondere bei der Beschreibenden Statistik wurde dabei Wert auf die Nutzung von realen Daten in gesellschaftlich bedeutsamen Kontexten gelegt. Die Beurteilende Statistik wurde dann mit einer Vorlesung und einer Präsenzübung zum Thema Signifikanztest mit der P-Wert-Methode begonnen. Die P-Wert-Methode geht zurück auf R. A. Fisher. In diesem Konzept wird die Nullhypothese verworfen, wenn die Wahrscheinlichkeit für das Auftreten eines Ergebnisses oder noch extremeren Ergebnissen unter Annahme der Nullhypothese sehr gering ist. Eine Alternativhypothese und ein Ab-

lehnungsbereich kommen in Fishers Konzept nicht vor.⁴ Im Anschluss an diese Einführung erhielten die Studierenden die im folgenden Abschnitt beschriebene Hausübung. Das Ziel der Behandlung der Aufgabe war es vor allem, dass die Studierenden üben, die Ergebnisse eines Signifikanztests zu interpretieren und dabei insbesondere vor dem Hintergrund der Frage „Was bedeutet ein signifikantes Ergebnis?“ die Grenzen dieser statistischen Methode kennenlernen.

3 „Die verlorenen Mädchen von Gorleben“

In der Hausübung wurde eine authentische Anwendung des Signifikanztests thematisiert, die im Frühjahr 2012 stark in der Presse diskutiert wurde. Anlass war eine Studie des Biostatistikers Hagen Scherb, der von einer statistisch signifikanten Abweichung im Geschlechterverhältnis Neugeborener rund um Gorleben sprach. Er untersuchte den Anteil der Jungen- bzw. Mädchengeburten nach dem Beginn der Einlagerung von Atommüll im Jahr 1996 in das seither politisch umstrittene Zwischenlager, das sich 2 km südlich von der niedersächsischen Gemeinde Gorleben befindet. Scherb und seine Kollegen werteten „die verlorenen Mädchen von Gorleben“ als Nachweis für eine stattgefundenen Strahlenbelastung (Scherb et al. 2012). Daraufhin überprüfte auch das Niedersächsische Landesgesundheitsamt (NLGA) die Daten sowie die daraus resultierenden Schlussfolgerungen, indem sie eine eigene Studie durchführten.

Liegt statistische Signifikanz vor?

Die Studierenden sollten im Rahmen ihrer Hausübung zunächst die Frage untersuchen, ob das Geschlechterverhältnis rund um Gorleben im Vergleich zu Deutschland tatsächlich statistisch signifikant verschoben ist.⁵ Dazu erhielten sie die vom NLGA 2011 veröffentlichten Geburtenzahlen:

	1991–1995	1996–2009	Summe
Jungen	3477	10040	13517
Mädchen	3421	9169	12590
Summe	6898	19209	26107

Tab. 1: Geburten in sämtlichen Gemeinden im 35 km-Umkreis um das Atommüllzwischenlager in Gorleben

Formulierung der Nullhypothese

Um nun einen Signifikanztest durchzuführen, muss zunächst eine zu testende Nullhypothese formuliert werden. Eine geeignete Nullhypothese wäre hier

„Die Wahrscheinlichkeit für eine Jungengeburt im 35 km-Umkreis um das Atommüllzwischenlager Gorleben ist nach der Einlagerung des Atommülls gleich der Wahrscheinlichkeit für eine Jungengeburt in Deutschland in demselben Zeitraum, d. h. $H_0: P(\text{Jungengeburt}) = 0,513$.“⁶

Statistisches Ergebnis und Interpretation

Der P-Wert, also die Wahrscheinlichkeit dafür, dass unter den 19209 Geburten im Zeitraum von 1996 bis 2009 10040 oder mehr Jungen geboren wurden, unter der Annahme, dass die Nullhypothese gilt, lässt sich mit Hilfe der Binomialverteilung berechnen: Sei X die Anzahl der geborenen Jungen. Mit $X \sim \text{Bin}(19209; 0,513)$ lässt sich der P-Wert wie folgt bestimmen:

$$P(X \geq 10040 \text{ unter der Annahme, dass } H_0 \text{ gilt}) =$$

$$\sum_{k=10040}^{19209} \binom{19209}{k} \cdot 0,513^k \cdot 0,487^{19209-k} \approx 0,37 \%$$

Alternativ kann der P-Wert auch durch Simulation mit dem TI-Nspire bzw. Fathom⁷ bestimmt werden (siehe Abb. 1).

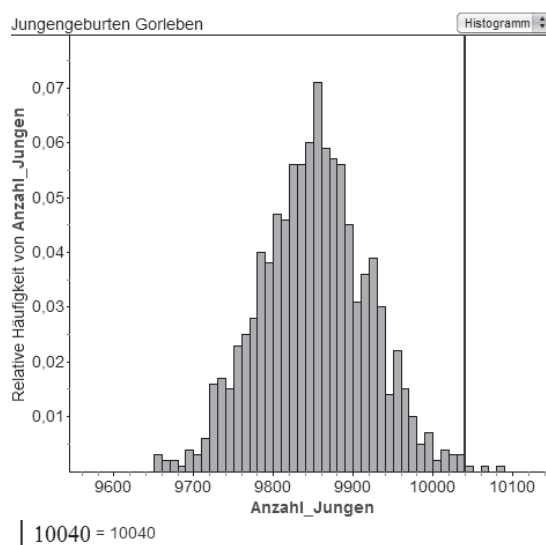


Abb. 1: Ergebnis einer mit Fathom erstellten 1000fachen Simulation von 19209 Geburten mit einer Wahrscheinlichkeit für eine Jungengeburt von 0,513.

Entscheidend ist nun aber die Frage, welche Schlüsse man aus diesem Ergebnis ziehen kann. Zwar wurde nachvollzogen, dass das Geschlechterverhältnis tatsächlich statistisch hochsignifikant verschoben ist. Allerdings weiß man nichts über die Ursachen für diese Verschiebung. Das Ergebnis des Signifikanztests gibt uns zunächst nur den Hinweis, dass es in der Umgebung von Gorleben etwas geben kann, dass zur Veränderung der Wahrscheinlichkeit für eine Jungengeburt seit 1996 geführt hat. Die genaue Ursache

ist aber nicht bekannt. Es konnten diesbezüglich auch keine Einflussfaktoren nachgewiesen werden. Insbesondere lag die messbare Strahlenbelastung durch das Atommüllzwischenlager weit unterhalb der natürlichen Strahlenbelastung (vgl. NLGA 2011; Krämer 2012). Letztendlich bleibt auch noch die Möglichkeit, dass die Wahrscheinlichkeit für eine Jungengeburt um Gorleben gar nicht von der Wahrscheinlichkeit für Deutschland abweicht, sondern dass es sich um eine zufällige Verschiebung im Geschlechterverhältnis handelt und vielleicht nur ein signifikantes Ergebnis unter vielen nicht-signifikanten ausgewählt wurde.

4 Studierendenbearbeitungen

Im Folgenden sollen nun die Schwierigkeiten bei der Formulierung der Nullhypothese anhand von sieben ausgewählten Studierendenlösungen untersucht werden:

- 1) „Der höhere Anteil an Jungengeburten hängt *nicht* mit der Nähe zum Atommüllzwischenlager zusammen.“
- 2) „Es werden *nicht* signifikant mehr Mädchen als Jungen geboren.“
- 3) „Die niedrige Mädchengeburtenrate liegt *nicht* an der Strahlung des radioaktiven Materials, sondern ist Zufall.“
- 4) „Es werden *mehr* Mädchen als Jungen geboren.“
- 5) „In Gorleben werden weniger Mädchen geboren.“
- 6) „Im Umkreis von Gorleben werden im Vergleich zu Deutschland weniger Mädchen geboren.“
- 7) „Es werden genauso viele Jungen geboren wie Mädchen.“

Verneinungen

Bei einem ersten Blick fallen zunächst die mehrmals vorkommenden *Verneinungen* auf (vgl. 1–4). Einen Grund dafür findet man womöglich bei Betrachtung der Beispiele, die zuvor in der Vorlesung behandelt wurden.

Eines davon ist das klassische Beispiel der sogenannten tea-tasting-lady. Sie behauptet am Geschmack zu erkennen, ob zuerst Tee oder Milch in die Tasse gegossen wurde und muss ihre Fähigkeit anschließend in einem Geschmackstest unter Beweis stellen. Möchte man nun eine Nullhypothese H_0 zum Testen formulieren, erhält man diese durch *Verneinung* der Behauptung: Die Lady kann die Reihenfolge *nicht* am Geschmack identifizieren, d. h. sie rät: $P(\text{„richtig identifizierte Tasse“}) = 0,5$.

Dasselbe Muster funktioniert auch bei dem zweiten in der Vorlesung behandelten Beispiel aus der Frankfurter Rundschau:

„Auch die alte Fußballweisheit, dass der Gefoulte nicht selbst den Elfmeter schießen sollte, haben Wissenschaftler unter die Lupe genommen – und widerlegt. Biometriker Oliver Kuß von der Martin-Luther-Universität Halle-Wittenberg hat alle Foulelfmeter der Bundesliga aus zwölf Jahren, exakt 835, untersucht. 102 davon wurden von den Gefoulten selbst geschossen und zu 73 Prozent verwandelt. Führten nicht-gefoulte Spieler den Strafstoß aus, ging der Ball in 75 Prozent der Schüsse ins Netz. ‚Dieser Unterschied liegt im Rahmen der zufälligen Schwankung und lässt somit nicht auf einen echten Effekt schließen‘, konstatiert Kuß. Seine Quintessenz: ‚König Fußball ist viel mehr vom Faktor Zufall bestimmt als viele Beteiligte glauben.‘“
(Mertens 2008)

Hier stellt die „alte Fußballweisheit“ die Behauptung dar: Die gefoulten Spieler haben eine schlechtere Trefferwahrscheinlichkeit als ihre Mitspieler. Durch *Verneinung* der Behauptung erhält man wiederum die zu testende Nullhypothese: Die gefoulten Spieler haben *keine* schlechtere Trefferwahrscheinlichkeit als ihre Mitspieler, d. h. ihre Trefferwahrscheinlichkeit ist gleich der der Mitspieler (oder besser).⁸ Aus diesen Beispielen könnten die Studierenden mit Schwierigkeiten beim Aufstellen der Nullhypothese also generalisiert haben: Wenn man die Behauptung bzw. Vermutung, die man über die Sachsituation besitzt, *verneint*, erhält man die zu testende Nullhypothese.

Überträgt man diese „Regel“ nun auf das Beispiel der „verlorenen Mädchen von Gorleben“, so kann dies funktionieren. Schließlich ist eine mögliche Behauptung: Im Umkreis von Gorleben ist die Wahrscheinlichkeit für eine Jungengeburt größer als in Deutschland. Dann erhält man durch *Verneinung* dieser Behauptung die zu testende Nullhypothese: Im Umkreis von Gorleben ist die Wahrscheinlichkeit für eine Jungengeburt *nicht* größer als in Deutschland, sondern gleich (oder kleiner). Andererseits ist eine andere mögliche Behauptung, die aus dem Sachkontext gewonnen werden kann: „Der höhere Anteil an Jungengeburten hängt mit der Nähe zum Atommüllzwischenlager zusammen.“ Diese Behauptung führt *verneint* zu der unter 1) formulierten Nullhypothese. Ebenso lassen sich entsprechende Behauptungen formulieren, die *verneint* zu den Nullhypothesen 2) bis 4) führen. All diesen *Verneinungsansätzen* ist aber gemein, dass sie keine statistisch zu testenden Hypothesen darstellen.

Man könnte nun schlussfolgern, dass diese Behauptungen nicht zur Sachsituation passen. Dies ist aber nicht

es macht im Sachkontext durchaus Sinn, diese vor dem Hintergrund der Theorie überprüfen zu wollen. Wenn es jetzt aber darum geht, eine Hypothese zu formulieren, die man mit Hilfe eines Signifikanztests überprüfen kann, muss zunächst eine geeignete Mathematisierung der Sachsituation vorgenommen werden. Dazu muss aber verstanden worden sein, was ein Signifikanztest in dieser Situation leisten kann. Dies scheint den Studierenden mit den oben beschriebenen Schwierigkeiten noch nicht klar gewesen zu sein.

Die Formulierungen 1) und 3) der Nullhypothesen stellen den Bezug zur Nähe zum Atommüllzwischenlager bzw. zur Strahlung des radioaktiven Materials her. Sie legen damit nahe, dass ein Schluss auf die Theorie, und damit auf die Ursache der statistisch signifikanten Abweichung möglich wäre. Anhand der vorliegenden Daten und bei jeder ökologischen Studie, bei der sich Einflussfaktoren nicht kontrollieren lassen, ist ein Schluss auf die Ursache jedoch nicht möglich.

Bei beiden Formulierungen ist aber klar, dass der Anteil an Jungen- bzw. Mädchengeburten feststeht und nicht überprüft werden soll. Dies ist bei den Formulierungen 4) bis 7) nicht der Fall. Die Studierenden formulieren die Nullhypothese dort so, als würde die Anzahl der geborenen Mädchen und Jungen rund um Gorleben und damit auch der Anteil der Mädchengeburten nicht feststehen. Es scheint für sie also fraglich zu sein, ob tatsächlich mehr oder weniger Jungen als Mädchen geboren werden. Dies steht aber fest. Es ist lediglich zu überprüfen, ob die beobachtete relative Häufigkeit nicht auch zustande kommen könnte, wenn rund um Gorleben dieselbe Wahrscheinlichkeit für eine Jungengeburt wie in Deutschland herrschen würde.

Ergänzende hilfreiche Kriterien

Ergänzend kann ein Ansatz aus der Beschreibenden Statistik von Arnold (2013) übertragen werden. Sie legte gemeinsam mit erfahrenen Lehrkräften Kriterien für gute Untersuchungsfragen im Rahmen von Datenanalysen fest. Übertragen auf den Signifikanztest, indem jeweils der Begriff „Frage“ durch „Nullhypothese“ ersetzt wurde und die Intention an den Zweck eines Signifikanztests angepasst wurden, ergeben sich folgende Kriterien, die Lernende als Hilfestellung beim Formulieren der zu testenden Nullhypothese nutzen können:

- Es lohnt sich, die Nullhypothese zu untersuchen.
- Die Nullhypothese kann mit den vorliegenden Daten untersucht werden.

- Die Intention ist klar: Es geht um die Bewertung von Daten bzw. Stichprobenergebnissen.
- Die interessierende Population wird benannt.
- Die interessierenden Merkmale/Zufallsgrößen werden benannt und es sind entsprechende Daten verfügbar.

Betrachtet man die Studierendenformulierungen der Nullhypothesen vor dem Hintergrund dieser Kriterien, sieht man, dass diese oftmals nicht erfüllt werden. Bei 1) und 3) handelt es sich um Hypothesen, die man mit den vorliegenden Daten nicht untersuchen kann, bei 4) bis 7) werden die interessierenden Merkmale falsch benannt und nur bei 6) wird die interessierende Population explizit angegeben.

6 Ausblick

Die im Rahmen der universitären Lehrveranstaltung gewonnenen Erkenntnisse lassen sich auf den Stochastikunterricht in der Sekundarstufe II übertragen, da die Schülerinnen und Schüler laut den Bildungsstandards der KMK lernen sollen, „Hypothesentests [zu] interpretieren und die Unsicherheit und Genauigkeit der Ergebnisse [zu] begründen“ (KMK 2012; S. 21).

Insgesamt sollte man dabei das diagnostische Potenzial der sprachlichen Formulierung der statistischen Nullhypothese der Lernenden stärker nutzen. Denn man erkennt bereits beim Formulieren der Nullhypothese, ob die Studierenden Fehlvorstellungen besitzen und somit die Grundidee des Signifikanztests noch nicht ausreichend verstanden haben. Dies wird auch an den Studierendenlösungen deutlich. An der eigentlichen Berechnung des P-Wertes kann dies hingegen nicht festgemacht werden: Ein Großteil der Studierenden, die die oben beschriebenen ungeeigneten Nullhypothesen formuliert haben, konnten den P-Wert mit Hilfe der Binomialverteilung bzw. mittels einer Simulation korrekt bestimmen, obwohl ihnen nicht klar war, zu welchem Zweck sie dies tun.

Problematisch ist es daher mit Blick auf den Oberstufenunterricht, dass viele Schulbücher sich auf die Durchführung von Hypothesentests zu vorgegebenen Hypothesen in meist eingekleideten Kontexten beschränken und dabei die Bestimmung von Ablehnungsbereichen in den Fokus stellen.

Einen guten Ansatz zum Finden geeigneter authentischer Anwendungen, anhand derer sich das Testen von Hypothesen thematisieren lässt, liefern Pressemeldungen, die sich mit den Ergebnissen von Signifikanztests beschäftigen. Nicht selten bekommt man in den Originalstudien geeignete Daten, um die

Ergebnisse statistisch zu überprüfen und auch die Angemessenheit der gegebenen Schlussfolgerungen zu hinterfragen. Kritisch ist dabei allerdings, dass im Schulunterricht häufig nur die Thematisierung von Binomialtests möglich ist, wenn Wahrscheinlichkeiten rechnerisch bestimmt werden sollen, da den Schülerinnen und Schülern kaum andere Wahrscheinlichkeitsverteilungen bekannt sind. In der Realität werden hingegen oftmals t-Tests oder andere Testverfahren verwendet. Ein Ausweg bietet diesbezüglich die Nutzung von Simulationen, durch die man auch Tests behandeln kann, die die Lernenden rechnerisch nicht bearbeiten könnten.¹¹

Im Sommersemester 2017 wird die Lehrveranstaltung zur „Elemente der Stochastik“ in neuer Konzeption umgesetzt. Dabei wird das Testen von Hypothesen an mehreren authentischen Anwendungen thematisiert, die sich allesamt auf Pressemeldungen beziehen. Der P-Wert lässt sich dabei jeweils durch eine Simulation bestimmen, wobei bei anderen zugrundeliegenden Verteilungen die Simulation aus der bei binomialverteilten Zufallsvariablen übertragen werden kann. Darüber hinaus steht die bewusste Unterscheidung zwischen der substantiellen Hypothese und dem Beitrag, den der Signifikanztest und das Verwerfen der statistischen Nullhypothese zur jeweiligen Sachfrage leisten können, im Fokus der Veranstaltung. Zu diesem Zweck werden die einzelnen Schritte des Vorgehens beim Signifikanztest mit Hilfe eines Modellierungskreislaufes reflektiert. Da sich alle Beispiele auf Pressemeldungen beziehen, kann immer auch die Angemessenheit von Schlussfolgerungen thematisiert werden und damit statistische Grundbildung gefördert werden.

Anmerkungen

- 1 „Authentisch“ ist hier im Sinne einer „objektiven Authentizität“ (vgl. Eichler (2015)) gemeint, d. h. die Übereinstimmung der behandelten Fragestellung mit einer Fragestellung in der Realität.
- 2 Der Artikel ist aus einem Vortrag anlässlich der Herbsttagung des Arbeitskreises Stochastik in Rostock am 01.10.2016 entstanden.
- 3 Die Lehrveranstaltung wurde von Prof. Dr. Katja Krüger unter Mitarbeit von Anna Schäfer und Thomas Wassong durchgeführt.
- 4 Eine Behandlung dieser Methode findet sich z. B. in Neue Wege (2012; S. 165 ff).
- 5 Alternativ ist auch eine zweite Herangehensweise möglich: Der Vergleich des Geschlechterverhältnisses vor und nach der Einlagerung des Atommülls. Eine Darstellung dieses Ansatzes findet sich im Online-Zusatzmaterial zu Schäfer (2016).

- 6 0,513 ergibt sich als arithmetisches Mittel der Anteile für eine Jungengeburt in Deutschland der Jahre 1996 bis 2009.
- 7 Eine ausführliche Darstellung der Simulation findet sich in Schäfer (2016). Die Simulationsdateien sind im zugehörigen Online-Zusatzmaterial erhältlich.
- 8 Es genügt hier, die Nullhypothese nur mit „gleich“ zu formulieren, da bei einer statistisch signifikanten Abweichung der P-Wert für alle $p > 0,75$ kleiner wäre als für $p = 0,75$, weil die Trefferanzahl unter Annahme dieser Nullhypothese als extremer einzuschätzen ist. Analog lässt sich dies auf die anderen Beispiele übertragen.
- 9 Chow (1996) bezieht sich eigentlich auf das in der Forschung häufig praktizierte und in Schulbüchern genutzte „null hypothesis significance-test procedure“ (S. ix), eine Vermischung des in diesem Artikel verwandten Signifikanztests mit der P-Wert-Methode und dem Hypothesentesten nach Neyman und Pearson. Obwohl die Studierenden vor der Bearbeitung das Konzept der Alternativhypothese (noch) nicht kennengelernt haben, lassen sich die Typen von Hypothesen übertragen.
- 10 Ein ähnlicher Modellierungskreislauf findet sich schon bei Schupp (1988), der aber vom „Modellieren“ statt „Mathematisieren“ und vom „Deduzieren“ statt „Verarbeiten“ spricht. Andere Ansätze zum Modellieren insb. in der Statistik und der Wahrscheinlichkeitsrechnung finden sich z. B. in Eichler und Vogel (2013).
- 11 Vorschläge dazu finden sich z. B. in Tintle u. a. (2015).

Literatur

- Arnold, P. M. (2013): Statistical investigative questions. An enquiry into posing and answering investigative questions from existing data. <https://researchspace.auckland.ac.nz/handle/2292/21305> (Zugriff: 31.01.17).
- Chow, S. L. (1996): Statistical significance. Rationale, validity and utility. London u. a.: Sage Publications.
- Eichler, A. (2015): Zur Authentizität realitätsorientierter Aufgaben im Mathematikunterricht. In: Kaiser, G.; Henn, H.-W. (Hrsg.): *Werner Blum und seine Beiträge zum Modellieren im Mathematikunterricht*. S. 105–118.
- Eichler, A.; Vogel, M. (2013): Daten- und Wahrscheinlichkeitsanalyse als Modellierung. In: Borromeo Ferri, R.; Greefrath, G.; Kaiser, G. (Hrsg.): *Mathematisches Modellieren für Schule und Hochschule*. Wiesbaden: Springer Spektrum.
- Engel, J. (2007): Daten im Mathematikunterricht: Wozu? Welche? Woher? In: *Der Mathematikunterricht* 53, S. 12–22.
- Engel, J. (2010): Anwendungsorientierter Mathematikunterricht: Von Daten zu Funktionen. Eine Einführung in die mathematische Modellbildung für Lehramtsstudierende. Berlin, Heidelberg: Springer.

- KMK (2012): Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife. Beschluss der Kultusministerkonferenz vom 18.10.2012. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf (Zugriff: 16.06.17).
- Krämer, W.: Die verlorenen Mädchen von Gorleben. Unstatistik vom 01.06.2012. <http://www.rwi-essen.de/unstatistik/7/> (Zugriff: 07.02.17).
- Gigerenzer, G. & Krauss, S. (2001): Statistisches Denken oder statistische Rituale? Was sollte man unterrichten? In: M. Borovcnik, J. Engel & D. Wickmann (Hrsg.): *Anregungen zum Stochastikunterricht: Die NCTM-Standards 2000, Klassische und Bayessche Sichtweise im Vergleich*. Hildesheim: Franzbecker, S. 53–62.
- Krüger, K. (2016): Statistische Grundbildung fördern. In: *mathematik lehren* 197, S. 2–7.
- Lergenmüller, A., Schmidt, G.; Krüger, K. (Hrsg.) (2012): *Mathematik. Neue Wege. Stochastik*. Braunschweig: Schroedel.
- Meehl, P. E. (1997): The Problem Is Epistemology, Not Statistics: Replace Significance Tests by Confidence Intervals and Quantify Accuracy of Risky Numerical Predictions. In: Harlow, L. L., Mulaik, S. A., Steiger, J. H. (Hrsg.): *What If There Were No Significance Tests?* Mahwah, NJ: Erlbaum 1997, S. 392–425.
- Mertens, M. (2008): Fußball macht klug. <http://www.fr.de/wissen/gehirnforschung-fussball-macht-klug-a-1184075> (Zugriff: 30.06.17).
- Nickerson, R.: Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy. In: *Psychological Methods* 5 (2), S. 241–301.
- Niedersächsisches Landesgesundheitsamt (2011): Veränderungen beim sekundären Geschlechterverhältnis in der Umgebung des Transportbehälterlagers Gorleben ab 1995. www.nlga.niedersachsen.de/download/60794 (Zugriff: 31.01.17).
- Schäfer, A. (2016): Weniger Mädchen in Gorleben?! Statistische Signifikanz verstehen. In: *mathematik lehren* 197, S. 40–44.
- Scherb, H.; Voigt, K.; Kusmierz, R. (2012): Verändertes Geschlechterverhältnis bei der Geburt nach Exposition gegenüber ionisierter Strahlung. Kernaussagen zur Präsentation und Pressekonferenz bei der Deutschen Umwelthilfe e. V. am 27.4.2012 in Berlin. https://www.helmholtz-muenchen.de/fileadmin/ICB/biostatistics_pdfs/scherb/GeschlechterVerhaeltnisKernaussagenfuerUmwelthilfe.pdf (Zugriff 07.02.17).
- Schupp, H. (1988): Anwendungsorientierter Mathematikunterricht in der Sekundarstufe I zwischen Tradition und neuen Impulsen. In: *Der Mathematikunterricht* 36 (6), S. 5–16.
- Tintle, N. u. a. (2015): *Introduction to statistical investigations*. Preliminary edition. New York: John Wiley & Sons Inc.

Anschrift der Verfasserin

Anna Schäfer
 Institut für Mathematik
 Universität Paderborn
 Warburger Straße 100
 33098 Paderborn
 annascha@math.upb.de