Bestell-Nr. 524214

MU

DER MATHEMATIK-UNTERRICHT

Blabla blabla blabla blabla p=0,4 blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla 800 blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla unter 40 % blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla höchstens 4 % blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla

Schickt die statistische Signifikanz in den Ruhestand!

Jahrgang 66 · Heft 4 · Juli 2020

Jahrgang 66 · Heft 4–2020

DER MATHEMATIK-UNTERRICHT

Beiträge zu seiner fachlichen und fachdidaktischen Gestaltung

Schwerpunkt **Schickt die statistische Signifikanz in den Ruhestand!**
Wolfgang Riemer

Adressen der Autoren

Norbert Henze
Henze@kit.edu

Thomas Hotz
thomas.hotz@tu-ilmenau.de

Wolfgang Riemer
w.riemer@arcor.de

Birgit Skorsetz
Birgit.Skorsetz@thillm.de

Reimund Vehling
vehling@icloud.com

BEITRÄGE

NORBERT HENZE, THOMAS HOTZ, WOLFGANG RIEMER, BIRGIT SKORSETZ, REIMUND VEHLING

Schickt die statistische Signifikanz in den Ruhestand!

1 Gute Botschaft

Der 30.05.2019 war ein guter Tag. Eine Meldung, die uns aus dem Herzen sprach: „Schickt die Signifikanz in den Ruhestand!“


Schickt die statistische Signifikanz in den Ruhestand!

Drei Statistiker fordern gemeinsam mit mehr als 800 weiteren Fachleuten, den p-Wert als Signifikanzkriterium aufzugeben: Er unterstelle zwei Kategorien von Ergebnissen, die es eigentlich nicht gibt.

Valentin Amrhein, Sander Greenland und Blake McShane

Abb. 1 Spektrum der Wissenschaft, 30.05.2019

„Wir“, das sind NORBERT HENZE vom KIT in Karlsruhe, der in seinem Standardwerk „Stochastik für Einsteiger“ zu Signifikanztests über 12 Auflagen hinweg notiert:

„Mit der Verfügbarkeit zahlreicher Statistik-Softwarepakete erfolgt das Testen statistischer Hypothesen in den empirischen Wissenschaften vielfach nur noch per Knopfdruck nach einem beinahe schon rituellen Schema.

Statistische Tests erfreuen sich u. a. deshalb einer ungebrochenen Beliebtheit, weil

- ihre Ergebnisse objektiv und exakt zu sein *scheinen*,
- *alle* von ihnen Gebrauch machen,
- der Nachweis der statistischen Signifikanz eines Resultates durch einen Test vielfach zum *Erwerb eines Doktortitels notwendig* ist“,

THOMAS HOTZ, der mit seinen Erfahrungen als Medizin- sowie amtlicher Statistiker an der TU Ilmenau statistische Beratungen für alle Fachbereiche anbietet und dort wie in seinen Vorlesungen sagt: „Egal wie das Testergebnis lautet, man lernt nichts daraus!“,

BIRGIT SKORSETZ, vom ThILLM, Bad Berka, die als Leiterin der Lehrplankommission den Schneid hatte, 2018 in Thüringen die Signifikanztests tatsächlich in den Ruhestand zu schicken und landesweit durch das Schätzen von Wahrscheinlichkeiten über Konfidenzintervalle abzulösen – und zwar ein Jahr VOR dem Artikel,

REIMUND VEHLING (Hannover) und WOLFGANG RIEMER (Köln), die seit zwei Jahrzehnten versuchten, als schulbuchschreibende Überzeugungslehrer das Leiden Hypothesen testender Schüler, Studenten, Referendare und Kollegen zu lindern.

2 Zweiseitige Signifikanztests

Das Gefährliche am Konzept der Signifikanztests ist die Verführung zum – nicht nur in der Schule – weit verbreiteten Schwarz-Weiß-Denken, zum Denken in den Kategorien Verwerfen/Akzeptieren, Wahr/Falsch. Wahrscheinlichkeiten sind aber vom Menschen gemachte Modelle der Wirklichkeit, und als Modelle sind sie nie ganz richtig und nur selten ganz falsch, sondern nur besser oder schlechter. Und statistische Testgrößen liefern nicht mehr als

behutsam zu interpretierende Indizien dafür, wie gut oder eben wie schlecht die Modelle die Wirklichkeit beschreiben.

Besonders viel Fingerspitzengefühl erfordert beim Einsatz von Signifikanztests die sachgerechte Beurteilung des Stichprobenumfangs.

Der Hauptsatz der beurteilenden Statistik lautet nämlich: „*Jede Hypothese kann auf jedem Signifikanzniveau verworfen werden, wenn man den Stichprobenumfang nur groß genug wählt.*“ Er ist allen Statistikern geläufig, wird aber aus naheliegenden Gründen in Schulbüchern ebenso „totgeschwiegen“ wie der fundamentale Unterschied zwischen Relevanz und Signifikanz.

Anwender erwarten von statistischen Tests mehr als die Aufforderung zu Fingerspitzengefühl oder behutsamem Interpretieren. Sie erwarten knackige Antworten auf griffige Fragen, und sie deuten „**keine signifikante Abweichung**“ folglich als **Bestätigung der Hypothese**.

Und das gilt nicht nur für Schüler, sondern – wie **Abb. 2** belegt – vor allem für praktizierende und publizierende Wissenschaftler.

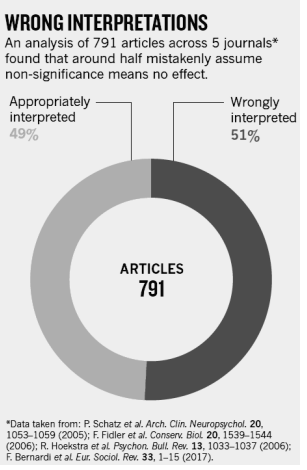

Abb. 2 AMRHEIN u. a. [2019]

Kasten 1: Binomialverteilung – zweiseitiger Signifikanztest (mit der 2σ-Regel)

Abb. 3

Ein Bleistift wird auf dem Tisch gerollt. Man vermutet, dass die beschriftete Seite mit einer Wahrscheinlichkeit p = 1/6 oben liegen bleibt.
JANA testet dies, indem sie ihren Bleistift 300-mal rollt. Dabei zählte sie 35 Versuche mit der bedruckten Seite nach oben. Spricht dieser Test gegen die obige Vermutung?
Eine Rechnung ergibt: 35 liegt außerhalb des 2σ-Intervalls [38; 62].
$p = 1/6$ wird daher „verworfen“. Anschließend folgt die gut gemeinte Warnung:
„Wir können nicht sicher sein, dass die Vermutung falsch ist. Denn in etwa 5 % aller Fälle können durchaus weniger als 38 oder mehr als 62 Treffer auftreten. Wenn wir also behaupten, aufgrund unseres Tests wäre die Vermutung $p = \frac{1}{6}$ widerlegt, so irren wir uns mit einer Wahrscheinlichkeit von etwa 5 %. Genauer gilt: $P(\text{Irrtum}) = P(X < 38) + P(X > 62) \approx 0{,}052$.“

https://www.schule-bw.de/faecher-und-schularten/mathematisch-naturwissenschaftliche-faecher/mathematik/unterrichtsmaterialien/sekundarstufe2/stochastik/testen (18.05.2020)

Aber auch **im Falle signifikanter Abweichungen** sind Fehldeutungen an der Tagesordnung, wie das schöne Experiment aus **Kasten 1** zeigt. JANA kann unter Benutzung der 2σ-Regel die Hypothese $p = 1/6$ auf dem ~5-%-Signifikanzniveau verwerfen und eine gute Antwort auf die gestellte Frage wäre:

„Das Testergebnis legt nahe, dass $p = 1/6$ kein sonderlich gutes Modell für JANAs Bleistift zu sein scheint." Nicht mehr und nicht weniger!

Wir nennen diese behutsame Aussage das „*Konzept des Bezweifelns*".

Die (gut gemeinte) Warnung in **Kasten 1** „Wir können nicht sicher sein, dass $p = 1/6$ falsch ist" führt in die Sackgasse. Tatsächlich kann man sogar absolut sicher sein, dass $p = 1/6$ nicht gilt. Um das sagen zu können, hätte JANA ihren Bleistift nach dem obigen Hauptsatz nicht einmal rollen müssen.

Und die verkürzte Aussage „Wenn wir $p = 1/6$ verwerfen, irren wir mit P(Irrtum) ~5 %" wird (nicht nur von Schülern) dahingehend gedeutet, dass wir beim Verwerfen zu ~95 % nicht irren, also richtigliegen. Und diese Deutung ist fatal, denn Tests machen keine Aussagen über Wahrscheinlichkeiten, mit denen Hypothesen gelten, sondern nur über Wahrscheinlichkeiten für (kritische) Ereignisse bei **unterstellter** Gültigkeit von Hypothesen, die eigentlich keinen Anwender interessieren.

Als Konsequenz fordert BUTH [2003, S. 30]: „Also gehören Signifikanztests in den Papierkorb, und zwar auf allen drei Ebenen: in der Mathematik, in den anwendenden Disziplinen und im Mathematikunterricht."

MOßBURGER [2014, S. 7] kommt zum Schluss „Wozu Signifikanztests? Was weiß man über die Gültigkeit einer Hypothese, wenn eine Stichprobe im Ablehnungsbereich A liegt? Nichts. Was folgt aus $P_H(A) \leq 5\,\%$? Nichts. Wozu soll man dann $P_H(A)$ berechnen? Um eine Meinung über eine Hypothese zu begründen, die mathematisch aber nicht begründet werden kann? Sind Signifikanztests in der Schule überhaupt sinnvoll?"

Wir beantworten diese Frage mit einem klaren NEIN!

Schülerinnen und Schüler, *die gelernt haben, Wahrscheinlichkeiten als Modelle der Wirklichkeit zu begreifen,* benutzen nach dem Bleistiftwürfeln ihren gesunden Menschenverstand und fragen nicht, wie wahrscheinlich $p = 1/6$ richtig oder falsch ist. Sie begeben sich auf die Suche nach einem besseren Modell. Die zufallsabhängige relative Häufigkeit $h = 38/300 \approx 12{,}7\,\%$ wäre ein besseres Modell … aber allein schon wegen der Zufälligkeit eben nicht das einzige. Und auf der Suche nach allen Wahrscheinlichkeiten, die man nicht bezweifeln muss, landet JANA mit ihrem Bleistift dann schneller, als man denkt, beim Konfidenzintervall [0,0845…; 0,1588] … und bei der Einsicht, dass es umso kleiner wird, je häufiger sie ihren Stift gerollt hat. Ganz im Gegensatz zum Hypothesentest wird hier der Einfluss des Stichprobenumfangs greifbar. Damit spiegeln Konfidenzintervalle eine Alltagserfahrung wider, über die schon Siebtklässler verfügen: „Je öfter, desto genauer."

Und wenn man die Länge eines Konfidenzintervalls als Messungenauigkeit deutet, dann schlagen Konfidenzintervalle auch eine Brücke zwischen Geometrie und Stochastik.

Eigentlich könnte an dieser Stelle unser Plädoyer gegen Signifikanztests und für Konfidenzintervalle enden, wären da nicht noch die einseitigen Signifikanztests.

Nach einem Blick auf die hier lauernden Tücken in Abschnitt 3 versuchen die beiden Schulpraktiker unter den Autoren in Abschnitt 4, denjenigen Kolleginnen und Kollegen Trost zu spenden, die in ihrem Bundesland immer noch testen müssen und noch nicht schätzen dürfen. Trost nicht mit warmen Worten, sondern mit (in Unterricht und Fachseminar) erprobten, konstruktiv-subversiven Tipps.

3 Einseitige Signifikanztests und der „gesunde Menschenverstand“

Wie oben ausgeführt, sind die Ergebnisse zweiseitiger Signifikanztests nicht leicht zu interpretieren. Da es aber nur eine Hypothese H_0: $p = p_0$ und nur eine Alternative H_1: $p \neq p_0$ gibt, wissen Schülerinnen und Schüler, was beim Lösen von Aufgaben zu tun ist, die Aufgaben sind machbar und Ansätze und Ergebnisse widersprechen nicht dem gesunden Menschenverstand, insbesondere dann nicht, wenn man die Aussage: Du kannst H_0 auf dem 10 %- (5 %-)Signifikanzniveau verwerfen im Sinne des „Konzepts des Bezweifelns“ vorsichtig und nur qualitativ interpretiert als: Begegne dem Modell p mit einer gewissen (einer etwas erhöhten) Skepsis … und suche ggf. nach besseren Modellen.

Bei einseitigen Signifikanztests ist die Situation vertrackter: Man hat neben der Wahrscheinlichkeit p_0 nun mehrere Nullhypothesen zur Auswahl: H_0: $p \geq p_0$, H_0: $p \leq p_0$, H_0: $p > p_0$, oder H_0: $p < p_0$. Welche Wahl „die richtige“ ist, hängt in Aufgaben (den Lehrplanvorgaben folgend) von versteckten „Interessen“ ab. Wenn man die Interessen vorgibt, lässt man die Nullhypothese suchen[1] – oder man gibt die Nullhypothese vor und lässt die versteckten Interessen suchen, wie in der folgenden Aufgabe „Ausflugsschiff“.

Ausflugsschiff (angelehnt an eine Abituraufgabe des IQB aus dem Jahr 2019)

Ein Unternehmen, das Fahrten mit einem Ausflugsschiff organisiert, bietet für 60 Plätze 64 Reservierungen an, weil **erfahrungsgemäß 10 %** der Reservierungen nicht erscheinen …
Das Unternehmen richtet ein Online-Portal zur Reservierung ein und **vermutet**, dass dadurch der Anteil p der Personen mit Reservierung, die zur jeweiligen Fahrt nicht erscheinen („no show“), **zunehmen** könnte. Als Grundlage für die Entscheidung darüber, ob pro Fahrt künftig mehr als 64 Reservierungen angenommen werden, soll die **Nullhypothese** „Die Wahrscheinlichkeit dafür, dass eine zufällig ausgewählte Person mit Reservierung nicht zur Fahrt erscheint, beträgt **höchstens 10 %**“ mithilfe einer Stichprobe von **200** Personen mit Reservierung auf einem **Signifikanzniveau von 5 %** getestet werden.
Vor der Durchführung des Tests wird festgelegt, die Anzahl der Reservierungen nur dann zu erhöhen, wenn die Nullhypothese abgelehnt werden musste.

- Ermitteln Sie für den beschriebenen Test die zugehörige Entscheidungsregel.
- Entscheiden Sie, ob bei der Wahl der Nullhypothese
  eher das Interesse, dass weniger Plätze frei bleiben sollen,
  oder das Interesse, dass nicht mehr Personen mit Reservierung abgewiesen werden müssen, im Vordergrund stand. Begründen Sie Ihre Entscheidung.

Hinter solchen zweiseitigen Testaufgaben verbirgt sich ein Muster mit stets wiederkehrenden Signalen, die es zu erkennen gilt.

[1] „Die Nullhypothese ist die Hypothese, an der man im eigenen Interesse tunlichst festhalten möchte, weil deren fälschliches Verwerfen schlimme Folgen hat.“

Aufgabenmuster
1. Wahrscheinlichkeit ($p_0 = 0{,}1$) herzaubern: Signal: „***erfahrungsgemäß 10 %***“
2. Konturen der Nullhypothese lüften: Signal:
   i) „***vermutet … zunehmen***“ signalisiert „rechtsseitiger Test“, denn die Nullhypothese soll das Gegenteil der Vermutung sein. Wer dieses Signal nicht entschlüsseln kann, bekommt hier zur Sicherheit (es geht um eine Abituraufgabe) zusätzlich
   ii) „***höchstens 10 %“*** und damit ist die Nullhypothese H_0: $p \leq 0{,}1$ vorgegeben.
3. Stichprobenumfang vorgegeben. Signal: ***natürliche Zahl*** (in Tabellen gelistet)
4. Signifikanzniveau vorgegeben. Signal: ***10 %, 5 %, 1 %***
5. Ablehnungsbereich ablesen/bestimmen
6. Unterstellte Interessen entschlüsseln

Lösung zu (6): Beim Ausflugsschiff möchte die Firma tunlichst auf H_0: $p \leq 0{,}1$ beharren und das Überbuchungskontingent nicht erhöhen, um sich vor Entschädigungszahlungen zu schützen, wenn gebuchte Plätze nicht verfügbar sind.

Tatsächlich will man hier eigentlich nachweisen, dass $p > 0{,}1$ gilt. Daher schlägt man in der Anwendung noch eine Volte, welche die Verwirrung zusätzlich erhöht: Man wählt die Nullhypothese als das **Gegenteil** dessen, was man belegen will, hier also H_0: $p \leq 0{,}1$, und fasst ein Verwerfen dieser Hypothese als Nachweis von $p > 0{,}1$ auf. Das ist aber nur sinnvoll, falls die Binomialverteilungsannahme zutrifft – und bei diesem Beispiel sagen Familienmitglieder oder Reisegruppenteilnehmer wohl kaum unabhängig voneinander ab.

Eine Urmutter aller Signifikanztest-Abituraufgaben „Reinlich und Sohn“ aus dem ersten NRW-Zentralabitur besitzt inzwischen einen zweifelhaften Kultstatus. Sie findet sich in STOYAN [2011], vgl. auch **Abb. 7**.

Das Problem bei einseitigen Signifikanztestaufgaben sind die durchgängig „an den Haaren herbeigezogenen“ Kontexte und die Verstöße gegen den gesunden Menschenverstand. Kein Statistiker, kein Unternehmen, das versucht, Gewinne mit Überbuchungsstrategien zu optimieren, käme auch nur im Ansatz auf die Idee, hier einseitige Signifikanztests zu bemühen und Interessen zu verstecken.

Der gesunde Menschenverstand würde stattdessen versuchen, die Verluste (durch nicht besetzte Plätze) und Entschädigungen (bei Überbuchungen) eher abzuschätzen, als funktional zu beschreiben (zu „modellieren“), und so entscheiden, dass die Gewinnerwartung möglichst groß wird.

Tatsächlich stiftet dieser Ansatz mit einer transparenten Bewertung von Fehlentscheidungen Einsicht und er lässt sich im Unterricht (in Spielsituationen auch handlungsorientiert) ausgezeichnet unterrichten [RIEMER 2020]. Er wird in den Bildungsstandards nicht erwähnt.

Abb. 4

4 Notfallambulanz

Bis Signifikanztests, insbesondere die einseitigen, flächendeckend durch Konfidenzintervalle abgelöst werden, bleibt nur die Suche nach Erster Hilfe mit dem Ziel, bleibende Schäden wie in **Abb. 4** abzuwenden. Zwei Ideen bieten wir an.

4.1 Erste Hilfe

Schüler goutieren Transparenz und Ehrlichkeit!

Man sollte als sachkundige Lehrperson das Standing haben, in aller Klarheit offenzulegen, dass die Aufgaben standardisierte Prüfungsformate bedienen, aber den gesunden Menschenverstand verachten.

Und wenn das klargestellt ist, kann man z. B. in arbeitsteiliger Partnerarbeit (durchaus mit subversiven Hintergedanken) Prüfungsaufgaben durch Markieren gängiger Signalworte dekontextualisieren (**Abb. 5, 6** nach einer Idee von M. VOGEL) und anschließend die dekontextualisierten durch Erfinden künstlicher Kontexte in Prüfungsaufgaben zurückverwandeln. Die dabei entstehende Heiterkeit wirkt befreiend. Der Unterricht gewinnt durch diese kritische Distanz eine ungeahnte emanzipatorische Tiefe. Und die Lernenden erleben, dass für Probleme mit zweiseitigen Signifikanztestaufgaben vielleicht eher die Aufgaben verantwortlich sind als mangelnde Intelligenz oder fehlender Fleiß.

Blabla blabla p=0,4 blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla 800 blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla unter 40 % blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla höchstens 4 % blabla blabla

Abb. 5 Abitur NRW (2007)

Blabla blabla 45,9 % blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla niedriger blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla 100 blabla blabla blabla blabla 2,5 % blabla blabla blabla blabla blabla blabla Entscheidungsregel blabla blabla ?

Abb. 6 Abitur Hessen (2016)

4.2 Zweite Hilfe

Die zweite Idee stammt von MAX, einem pfiffigen Schüler vom HMG, Köln, der lieber über Probleme nachdenkt, als stumpfsinnig kritische Werte in (analogen oder digitalen) Tabellen aufzusuchen. Er nutzt zum Aufgabenlösen, vor allem aber zum Aufgabenverstehen, die GeoGebra Datei „signifikanztesttool.ggb“, **Abb. 7**. Nach Eingabe von Nullhypothesenwahrscheinlichkeit p_0, Signifikanzniveau α_0 und Stichprobenumfang n liefert sie den kritischen Bereich, grafisch unterstützt durch Einblenden der Binomialverteilung.

Wer lieber die Normalverteilung nutzt (hier aus Gründen der Übersichtlichkeit gespiegelt) wird dadurch auch „bedient“ – und erkennt, dass beide Berechnungsvarianten (fast immer) zum gleichen Ergebnis führen.

Und hier die Idee von MAX:

„Bei *einseitigen* Testaufgaben zum *5 %*-Signifikanzniveau teste ich ohne Rücksicht auf irgendwelche Interessen und Hintergedanken immer *zweiseitig* auf dem *10 %*-Niveau.“[2] Das liefert exakt die in den einseitigen Aufgaben benötigten 5 %-Verwerfungsbereiche, die in den „Zipfeln“ links und rechts vom 90 %-Prognoseintervall zu p_0 liegen.

1. Wenn das Testergebnis X im linken Zipfel liegt, gehen beide von $p < p_0$ bzw. $p \leq p_0$ aus.
2. Wenn das Testergebnis X im rechten Zipfel liegt, gehen beide von $p > p_0$ bzw. $p \geq p_0$ aus.

[2] MAX: „Sind einseitige Tests u. a. deswegen so populär, weil sie gestatten, bei gleicher Datenlage das publikationsentscheidende Signifikanzniveau zu halbieren?“

3. Nur dann, wenn das Testergebnis **im Prognoseintervall** liegt, konstruieren die Aufgabensteller Konflikte: Wer rechtsseitig testet, geht von $p \leq p_0$, wer linksseitig testet, von $p \geq p_0$ aus und rechtfertigt damit sein im Aufgabentext unterlegtes Handlungsmuster. Beim Ausflugsschiff soll rechtsseitig getestet, also dann von $p \leq p_0 = 0{,}1$ (kein erhöhtes no show) ausgegangen werden, um die Überbuchungsquote nicht erhöhen und die evtl. häufiger anfallenden Entschädigungen nicht zahlen zu müssen.

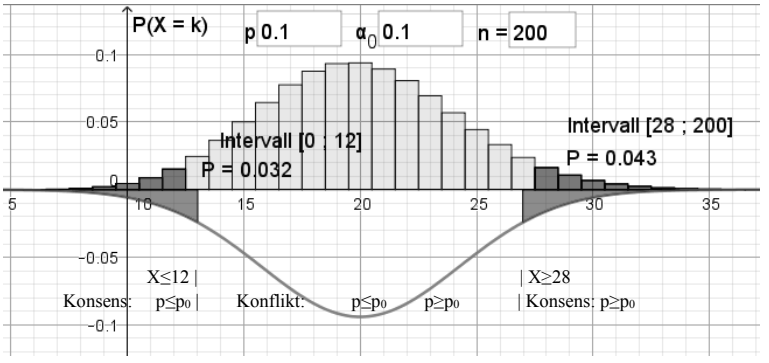

Abb. 7 signifikanztesttool.ggb

Präziser als MAX mit seiner GeoGebra Datei kann man die Rezeptur hinter einseitigen konstruierten Signifikanztestaufgaben nicht auf den Punkt bringen:

Wähle ein Testergebnis im Prognoseintervall von p_0 und konstruiere aus dem Konflikt eine Aufgabe, die mit gesundem Menschenverstand nur noch wenig zu tun hat.

Und damit sind wir auch auf dem Umweg über einseitige Signifikanztests wieder bei Prognoseintervallen gelandet: Man sollte sie nicht nutzen, um Interessen zu verstecken, sondern um sich auf die Suche nach der Wahrheit begeben. Die absolute Wahrheit, die „wahre Wahrscheinlichkeit", wird man nie finden. Begnügen wir uns mit vertrauenswürdigen Modellen, mit Konfidenzintervallen eben!

Literatur

[1] AMRHEIN, V. GREENLAND, S. MCSHANE, B. [2019]: Retire statistical significance. Nature 567, S. 305–307. 2019.
[2] BUTH, M. [2002]: Anmerkungen zum Testen von Hypothesen. SiS 2002/2, S. 27–29.
[3] HENZE, N. [2018]: Stochastik für Einsteiger [2018]. Berlin, Springer, 12. Auflage.
[4] MOßBURGER, M. [2014]: Unklare Begriffe und Wunschdenken bei Signifikanztests. SiS 2014/1 S. 2–8.
[5] RIEMER, W., SEEBACH, G. [2011]: Bleistiftrollen – Beurteilende Statistik im Federmäppchen. In R. KAENDERS, R. SCHMIDT (Hrsg.): Mit GeoGebra mehr Mathematik verstehen. 2. Aufl. 2014. Springer Spektrum S. 91–106.
[6] RIEMER, W. [2020]: Signifikanztests und das Risiko falscher Entscheidungen: Gewinn besiegt Wahrheit. ml 3/2020.
[7] STOYAN, D. [2011]: Statistische Tests in Gymnasiallehrbüchern. SiS 2011/1, S. 28–32.

Der Artikel aus **Abb. 1**, das Original AMRHEIN [2019] und signifikanztesttool.ggb sind erhältlich bei w.riemer@arcor.de.