Wie man das Testen von Hypothesen einführen sollte

STEFAN KRAUSS UND CHRISTOPH WASSNER

Zusammenfassung

Der vorliegende Artikel ist eine Ausarbeitung eines Vortrags, der bei der GDM-Tagung 2000 in Potsdam gehalten wurde. Eine Kurzfassung dieses Artikels findet sich im Tagungsband „Beiträge zum Mathematikunterricht 2000" unter dem Titel „Probleme bei der Interpretation signifikanter Testergebnisse". Wir möchten in diesem Artikel auf eine Lücke im Unterricht der Beurteilenden Statistik hinweisen. Schüler lernen Signifikanztests heute üblicherweise als eine Entscheidungsregel kennen: Ist ein Ergebnis signifikant, wird die Hypothese H_0 verworfen. Anhand einer empirischen Untersuchung zeigen wir, dass das zugrundeliegende Konzept der Signifikanz dabei im Dunkeln bleibt und einem signifikanten Testergebnis falsche Bedeutungen zugemessen werden. Wir machen einen didaktischen Vorschlag, wie man ein wirkliches Verständnis des Konzepts der Signifikanz erreichen kann

Brauchen wir überhaupt Signifikanztests in der Schule?

In den meisten Bundesländern lernen die Schüler – meist als Anhängsel an den Stochastikunterricht – Hypothesen mit Hilfe von Signifikanztests zu prüfen. Sogar dieser kleine Ausschnitt aus der Welt der Inferenzstatistik steht derzeit auf dem Prüfstand. Deschauer (1999) schreibt z.B.: „Ich denke doch, dass man in den frühen achtziger Jahren die Chancen eines anwendungsorientierten Mathematikunterrichts zu optimistisch eingeschätzt und die Stochastik-Leistungskurse etwas zu bombastisch ausgebaut hat. Welcher fachliche Aufwand muss betrieben werden, damit man zur Beurteilenden Statistik gelangt! Welcher Leistungskurs-Absolvent ohne nachfolgendes Mathematikstudium kommt einmal in die Situation, Hypothesen testen zu müssen!"

Nun, die Antwort hierauf ist einfach: Jeder Soziologie-, Psychologie-, Pädagogik-, Wirtschaftsstudent usw. hat sich mit statistischen Verfahren zu befassen. Statistik ist das meistunterrichtete Fach an deutschen Universitäten[1]. In nahezu allen Sozialwissenschaften (wie zum Beispiel Soziologie oder Psychologie) muss im Grundstudium ein Schein in Statistik, im Vordiplom eine Klausur in Statistik und in der Diplomarbeit die Fähigkeit zur Anwendung von Verfahren der Beurteilenden Statistik nachgewiesen werden. Diepgen wies bereits 1985(a) darauf hin, dass statistische Standardverfahren wie Signifikanztests in den sozialwissenschaftlichen Disziplinen (Diepgen nennt sie „Humanwissenschaften") aber leider oft missinterpretiert werden („unverstandene Rezeptbuchstatistik"). In der Psychologie gibt es mittlerweile eine fast schon unüberschaubare Fülle von Literatur, die von (zumeist unabsichtlichem) Missbrauch statistischer Methoden und den daraus resultierenden gravierenden Folgen berichtet (einen aktuellen Überblick über die sog. „Signifikanztestdebatte" in der Psychologie gibt z.B. Nickerson, 2000). Wir glauben, dass man die Signifikanztests in der Schule nicht etwa abschaffen, sondern der derzeitigen Hochschulrealität Rechnung tragen sollte und diese Tests in der Schule besonders verständlich unterrichten sollte. Auch Diepgen forderte bereits 1985(a): „Den gängigen Irrationalismen statistischer Alltagspraxis vorzubeugen, sollte daher meines Erachtens wichtige Aufgabe eines wissenschaftspropädeutischen und kritischen Statistikunterrichtes auf der Schule sein." (S.32)

Wir glauben, dass gerade durch die Art und Weise, wie derzeit Signifikanztests in der Schule unterrichtet werden, diese Irrationalismen erst entstehen können. Nach einem üblichen Curriculum können Schüler zwar meist – nach erheblichen Anlaufproblemen – die für das Testen von Hypothesen erforderlichen Berechnungen durchführen und Prüfgrößen anhand eines Tabellenwerks auf „Signifikanz" untersuchen. Dieser Vorgang hat mit dem wirklichen Verständnis des zugrunde liegenden Konzepts jedoch wenig zu tun, sondern gleicht eher einem mechanisierten Ritual. Die Folgerungen, die derzeit aus der TIMS-Studie gezogen werden, sowie die Forderungen der NCTM-Standards 2000 legen aber nahe, dass den Schülern statt mechanischem Einsetzen eher Verständnis für die Bedeutung der verwendeten Verfahren vermittelt werden sollte.

Anhand einer empirischen Untersuchung illustrieren wir im Folgenden, worin die Verständnisprobleme im Zusammenhang mit Signifikanztests liegen.

[1] In den Sozialwissenschaften steht „Statistik" dabei als Synonym für Stochastik.

1 Eine empirische Untersuchung: Was bedeutet eigentlich „signifikant"?

Wir legten Psychologiestudenten[2] der Freien Universität Berlin folgenden Fragebogen vor (der erstmals 1986 von Oakes verwendet wurde):

Stellen Sie sich vor, Sie prüfen mit Hilfe eines t-Tests[3] statistisch, ob der zwischen zwei Stichproben gefundene Unterschied signifikant ist. Es stellt sich heraus, dass der Unterschied auf dem 1%-Niveau signifikant ist. Welche der folgenden Aussagen lassen sich nun aus dieser Tatsache folgern? (Es können dabei alle oder auch keine Aussage zutreffen)

1) Es ist eindeutig bewiesen, dass die Nullhypothese falsch ist.

2) Die Wahrscheinlichkeit des Zutreffens der Nullhypothese ist gefunden worden.

3) Es ist eindeutig bewiesen, dass die Alternativhypothese wahr ist.

4) Man kann nun die Wahrscheinlichkeit ableiten, dass die Alternativhypothese richtig ist.

5) Entscheidet man sich nun, die Nullhypothese 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.

Alle diese Aussagen sind Illusionen über die Bedeutung eines signifikanten Testergebnisses (für eine Diskussion der einzelnen Aussagen siehe z.B. Gigerenzer, 1993a). Dennoch entschieden sich alle der von uns befragten Studenten für eine oder mehrere dieser Aussagen. Pikanterweise fanden wir in einer zweiten Untersuchung heraus, dass sich sogar Statistiklehrende an der Universität von diesen Aussagen verleiten lassen: Selbst in einer solchen Stichprobe glaubten noch 86% der Befragten fäschlicherweise an die Richtigkeit einer oder mehrerer der Aussagen (Haller und Krauss, eingereicht).

Warum sind diese Aussagen alle falsch? Der Begriff „Signifikanz" impliziert immer „eine Welt", in der H_0 gültig ist: Man betrachtet die *Wahrscheinlichkeit seiner Daten* unter der Annahme, H_0 sei richtig. Diese Wahrscheinlichkeit – in Symbolen $P(D \mid H_0)$[4] – wird auch „P-Wert" genannt. Wie vom Erfinder der Signifikanztests, Sir Ronald Fisher, gefordert, wird durch einen P-Wert ein *Signifikanzniveau der Daten* angegeben. Bedauerlicherweise spielt das Konzept der P-Werte bei der Einführung von Signifikanztests im Stochastikunterricht in der Schule keine Rolle[5]. Im Gegensatz dazu ist das uns aus Schulbüchern vertraute Signifikanzniveau keine Eigenschaft der Daten, sondern eine vorher festgelegte Eigenschaft des Tests: Wir interessieren uns im Stochastikunterricht nicht für den numerischen P-Wert, sondern nur dafür, ob dieser größer oder kleiner als ein *vorgegebenes Signifikanzniveau* ist (meist 5% oder 1%). Die Schüler haben so lediglich den empirischen Wert einer Prüfgröße mit einem kritischen Wert in einem Tabellenwerk zu vergleichen. Diesen Größenvergleich, der dann die Entscheidung bestimmt, ob H_0 angenommen oder verworfen wird („Entscheidungsregel"), kann ein Schüler mechanisch durchführen, ohne Verständnis für die eigentliche Bedeutung eines signifikanten Testergebnisses entwickelt zu haben. Die eigentliche Grundlage des Signifikanztestens, nämlich eine *Wahrscheinlichkeitsbetrachtung von Daten*, wird nicht durch das übliche Signifikanzniveau *eines Tests*, sondern nur durch die Betrachtung des Signifikanzniveaus *der Daten* (P-Wert) erhellt. Hätte man den P-Wert eingeführt, ließe sich der Begriff der Signifikanz mit folgendem kurzen Satz definieren: Ein Testergebnis ist genau dann auf dem 1%-Niveau signifikant, wenn der zum Testergebnis gehörige P-Wert kleiner als 1% ist. Dies bedeutet

[2] Wir möchten an dieser Stelle noch einmal darauf hinweisen, dass im Psychologiestudium der ersten beiden Semester, im Vordiplom und bei der Diplomarbeit eine intensive Beschäftigung mit Signifikanztests verlangt wird. Eine gewisse Routine bei der Durchführung von Signifikanztests kann also bei allen Probanden vorausgesetzt werden.

[3] Der t-Test zur Überprüfung von Mittelwertsunterschieden ist der von Psychologiestudenten am häufigsten benutzte Test. Selbstverständlich gelten unsere Betrachtungen aber für jeden beliebigen Signifikanztest.

[4] Genau steht D dabei für die „vorliegenden oder noch unwahrscheinlichere Daten". Die üblicherweise nicht zu findende Schreibweise als bedingte Wahrscheinlichkeit – $P(D \mid H_0)$ – ist ein wesentlicher Baustein unseres folgenden didaktischen Vorschlags. Es ist jedoch zu beachten, dass diese Schreibweise im Fall einer zusammengesetzten H_0 nicht existiert: In diesem Falle müsste erst das Maximum der Ablehnwahrscheinlichkeiten über alle Parameter im Hypothesenbereich bestimmt werden. Erst wenn H_0 mit genau diesem Parameter spezifiziert wird, ließe sich ein P-Wert bestimmen, der sich dann allerdings nicht in der Form $P(D \mid H_0)$ ausdrücken lässt (mehr dazu siehe Fußnote 9).

[5] Dies trifft auch auf die meisten Hochschullehrbücher in Wahrscheinlichkeitsrechnung zu, die für Fachmathematiker verfasst sind. Werden P-Werte tatsächlich erwähnt, so meist nur beiläufig (siehe z.B. Krengel, 1990).

aber: *Die Wahrscheinlichkeit der gefundenen (oder noch unwahrscheinlicherer) Daten* ist kleiner als 1%, wenn tatsächlich die Nullhypothese stimmen würde. Aber P-Werte haben noch weitere Vorteile: Gibt man statt der üblichen Mitteilung über „Signifikanz" eines Testergebnisses gleich den P-Wert als Endergebnis an, gewinnt man an Genauigkeit. Ein Leser weiß nun nämlich nicht nur, ob die Wahrscheinlichkeit der Daten unter H_0 größer oder kleiner als 5% ist, mit einem P-Wert weiß er sogar, wie groß dieser Wert tatsächlich ist. Der Fehler erster Art kann nun genau angegeben werden, und es gibt keinen Grund mehr, ihn mit einer willkürlich festgesetzten Schranke zu vergleichen; man benötigt im Grunde genommen das Signifikanzkonzept gar nicht mehr. Ob man nun aber die konventionellen Schranken (1% oder 5%) als Entscheidungskriterien beibehalten möchte oder nicht, wir denken, dass ein P-Wert viel zur Erhellung der zugrunde liegenden Test-Idee beitragen kann.

Der didaktisch entscheidende Punkt ist aber der Folgende: Aus der Schreibweise eines P-Wertes als bedingte Wahrscheinlichkeit wird klar, dass ein Signifikanztest zwar Aussagen über die *Wahrscheinlichkeit von Daten*, aber *nicht über die Wahrscheinlichkeit von Hypothesen* machen kann. Genau diese Fehler geschehen aber in den Aussagen 1–5: In den Aussagen 2 und 4 ist explizit von „Wahrscheinlichkeit von Hypothesen" die Rede. Die Aussagen 1 und 3 behaupten, mit einem signifikanten Testergebnis könne man Hypothesen sogar „beweisen". Da dies nichts anderes bedeutet als einer Hypothese die Wahrscheinlichkeit 1 zuzuschreiben, ist also auch hier von Wahrscheinlichkeiten von Hypothesen die Rede. Aussage 5 sagt, dass man die Wahrscheinlichkeit kennen würde, bei Verwerfen der H_0 einen Fehler gemacht zu haben. Auch dies würde bedeuten, dass man die Wahrscheinlichkeit der Nullhypothese kennt (die Aussagen 2 und 5 sind somit gleichbedeutend). Aussage 6 schließlich ist die sogenannte „replication fallacy" (mehr dazu z.B. bei Falk und Greenbaum, 1995). Wir werden im Folgenden noch sehen, warum auch diese Aussage falsch ist.

Die Schwierigkeiten bei der Interpretation eines signifikanten Testergebnisses sind weit verbreitet und werden in der sozialwissenschaftlichen Literatur oft berichtet. Auf dem Gebiet der Psychologie gibt es neben unserer Untersuchung (Haller & Krauss, eingereicht) auch noch die experimentellen Untersuchungen von Oakes (1986) und von Falk und Greenbaum[6] (1995). Gigerenzer (1993b) führt sogar Statistiklehrbücher an, in denen die Autoren selbst zur Erläuterung des Begriffes Signifikanz explizit auf Aussagen der Bauart 1–6 zurückgreifen.

Auch die Literatur zur Schulstochastik beschäftigt sich schon länger mit den Problemen bei der Durchführung und der Interpretation von Signifikanztests. Die von Ian Birnbaum (1982) berichteten Schülerzitate zum Begriff „Signifikanz" zeigen z.B., dass dieses Thema auch für die Didaktik der Gymnasialstochastik relevant ist. Möchte man Schüler vor den Folgen einer Überbewertung eines signifikanten Testergebnisses warnen, sind v.a. die Artikel von Diepgen (1985b) und von Waldmüller (1998) zu empfehlen. Um das Verständnis für Signifikanztests zu fördern, plädiert Stein (1993) für eine Präzisierung der Nullhypothese, während Gigerenzer und Krauss (2001) empfehlen, einen geschichtlichen Überblick über die Entstehung der "Signifikanzkonfusion" zu geben. Anknüpfend an diese und weitere bisherige didaktische Bemühungen, das Testen von Hypothesen verständlicher darzustellen (z.B. Riemer, 1985; Diepgen, 1985a; Buth, 1991; Buth, 1993), wollen wir hier vorschlagen, die eigentliche Idee des Signifikanztestens durch einen einfachen *Notationenvergleich eines P-Wertes mit der Formel von Bayes* transparent zu machen. Unser Ansatz richtet sich dabei gezielt auf die Auslöschung von Missinterpretationen der oben geschilderten Art. Anders als z.B. Buth (1991), der in seinem Ansatz aus didaktischen Gründen auf Bayesianisches Testen als Alternative zum Signifikanztesten zurückgreift, wollen wir die Formel von Bayes sozusagen als „Kontrastmittel" zum Signifikanztesten verwenden. Welche Schritte notwendig sind, bevor man mit Hilfe der Formel von Bayes verdeutlichen kann, was ein signifikantes Ergebnis alles *nicht* bedeutet, erläutern wir im Folgenden.

2 Ein didaktischer Ansatz: Was lässt sich gegen falsche Vorstellungen von Signifikanz tun?

Lassen Sie uns im Lehrplan einen Schritt zurückgehen: Bei der Beschäftigung mit bedingten Wahrscheinlichkeiten wird – wenn überhaupt – auch die Formel von Bayes unterrichtet. Dass diese Formel jedoch nicht nur eine Rechenregel ist, sondern eine zentrale Rolle in der Inferenzstatistik (also beim Beurteilen von Hypothesen) spielt, bleibt im Rah-

[6] Falk und Greenbaum konnten auch zeigen, dass selbst eine intensive Beschäftigung mit falschen Vorstellungen von Signifikanz Studenten nicht davon abhält, später dennoch an (geringfügig abweichende) Fehldeutungen zu glauben.

men des üblichen Stochastikunterrichts meist völlig im Dunkeln[7]. Wenn man nämlich die in Schulbüchern üblicherweise in der Formel auftauchenden allgemeinen Ereignisse A und B als Hypothesen H bzw. als Daten D interpretiert[8], erlaubt die Formel – im Gegensatz zu Signifikanztests –, tatsächlich Aussagen über die (bedingte) Wahrscheinlichkeit von Hypothesen zu treffen:

$$P(H \mid D) = \frac{P(D \mid H) \cdot P(H)}{P(D \mid H) \cdot P(H) + P(D \mid \bar{H}) \cdot P(\bar{H})}$$

In dieser Form begründet die Formel von Bayes einen mit dem klassischen Signifikanztest konkurrierenden Zugang zum Hypothesentesten. Der didaktische Kniff zur besseren Durchdringung der Signifikanztests besteht nun darin, *beide* Testparadigmen (klassisches und Bayessches) in Abhängigkeit von Daten und Hypothesen als bedingte Wahrscheinlichkeiten auszudrücken und vergleichend gegenüberzustellen:

(a) $P(H \mid D)$ liegt Bayesschen Testverfahren zugrunde

(b) $P(D \mid H_0)$ liegt Signifikanztests zugrunde

Diese Einsicht dürfte bei Schülern das Verständnis um die Prinzipien inferenzstatistischer Verfahren entscheidend voranbringen. Folgende Sachverhalte werden aus dieser Gegenüberstellung unmittelbar klar[9]:

• Es wird deutlich, dass der Satz von Bayes und Signifikanztests nicht zwei völlig voneinander losgelöste Konzepte sind. Vielmehr stehen *beide* Konzepte mit dem Begriff des Hypothesentestens in Verbindung. Weiterhin lassen sich *beide* Konzepte mit Hilfe bedingter Wahrscheinlichkeiten ausdrücken[10].

• Der Vergleich von (a) und (b) zeigt darüber hinaus, dass es zwei grundsätzlich verschiedene Arten Beurteilender Statistik gibt: Man kann die *Wahrscheinlichkeit der Hypothese bei vorgegebenen Daten* berechnen oder man kann die *Wahrscheinlichkeit der Daten bei gegebener Nullhypothese* betrachten. Insofern kann der Vergleich von (a) und (b) auch als ein erster vorsichtiger Schritt in Richtung des von den NCTM-Standards geforderten Methodenpluralismus im Mathematikunterricht gesehen werden.

Wir behaupten, dass man genau mit dieser Gegenüberstellung automatisch auch Klarheit über den Wahrheitsgehalt von Aussagen der Art 1–6 schafft: Vergleicht man die Notationen (a) und (b), sieht man unmittelbar, dass man nur mit einem dieser beiden Zweige der Beurteilenden Statistik Wahrscheinlichkeiten von Hypothesen berechnen kann, nämlich mit der Bayesstatistik. Da aber die Aussagen 1–5 alle behaupten, mit einem Signifikanztest könne etwas über die Wahrscheinlichkeit von Hypothesen heraus-gefunden werden, müssen sie automatisch falsch sein. Aus (b) wird weiterhin klar, dass jede Aussage, die die Bedeutung eines signifikanten Testergebnisses paraphrasieren soll, in irgendeiner Form zum Vorliegen von H_0 Stellung nehmen muss (z.B. „..., unter der Vorraussetzung, dass H_0 gilt"). Auch diese Bedingung ist bei allen Aussagen (diesmal inklusive Aussage 6) nicht erfüllt.

Was muss nun im Unterricht geändert werden, um zu diesen Einsichten zu gelangen? Der Satz von Bayes und das Testen von Hypothesen sind in einem herkömmlichen Curriculum durch eine Kluft getrennt, als ob es sich hierbei um zwei völlig verschiedene Dinge handelte: Der Satz von Bayes wird ohne den Begriff der Hypothese gelehrt und das Konzept der Signifikanz ohne Rückgriff auf bedingte Wahrscheinlichkeiten. Der rote Faden wird erst wieder sichtbar, wenn beide Gebiete miteinander verbunden werden. Um dies zu erreichen, muss der Satz von Bayes als Inferenzmöglichkeit (A und B können auch Hypothesen H und Daten D sein!) identifiziert und das Konzept der Signifikanz als bedingte Wahrscheinlichkeit dargestellt werden. Lassen Sie uns abschließend die im Unterricht nötigen Änderungen tabellarisch zusammenfassen:

[7] Gymnasiallehrer sind oft verwundert über die intensive Beschäftigung mit dieser ihrer Meinung nach eher unwichtigen Formel. Diese Verwunderung hat jedoch zwei Seiten: Erzählt man Hochschulstatistikern, dass im Wahrscheinlichkeitsunterricht am Gymnasium der Satz von Bayes nur ein Schattendasein fristet, reagieren diese wiederum mit Ungläubigkeit.

[8] Dinges (1978) fordert *grundsätzlich*, dass der Hypothesenbegriff im Stochastikunterricht stärker berücksichtigt werden sollte.

[9] Es sei an dieser Stelle noch einmal darauf hingewiesen, dass sich der P-Wert nur für Punkthypothesen als $P(D \mid H_0)$ schreiben lässt (siehe auch Fußnote 4). Die didaktischen Vorteile dieser intuitiven Darstellung sind unserer Meinung nach aber so gewichtig, dass man die Schreibweise deshalb nicht aufgeben, sondern einfach auf diesen Sonderfall hinweisen sollte.

[10] Es ist grundsätzlich möglich, *alle* Wahrscheinlichkeiten als bedingte Wahrscheinlichkeiten (z.B. gegeben Ω) auszudrücken. Bei Lindley (1965, S.6) findet sich sogar die Axiomatisierung der Wahrscheinlichkeitstheorie mit bedingten Wahrscheinlichkeiten.

Folgerungen für das Curriculum in Stochastik

Übliches Verfahren

Zunächst ...	Danach ...
... werden bedingte Wahrscheinlichkeiten und der Satz von Bayes mit allgemeinen Ereignissen A und B eingeführt – ohne die Berücksichtigung des Begriffs der Hypothese H.	... werden davon losgelöst Signifikanztests und der Begriff der Hypothese H eingeführt, und zwar unabhängig von bedingten Wahrscheinlichkeiten.

Änderungsvorschlag:

Erforderliches Verfahren

Zunächst ...	Danach ...
... sollten bedingte Wahrscheinlichkeiten und der Satz von Bayes *inklusive des Begriffs der Hypothese H* eingeführt werden.	... sollten Signifikanztests auch unter Zuhilfenahme *bedingter Wahrscheinlichkeiten* dargestellt und über den Vergleich von (a) und (b) der *Formel von Bayes gegenübergestellt* werden.

3 Einige Schlussbemerkungen

Signifikanztests sind bei weitem nicht das einzige stochastische Verfahren, das zu Missinterpretationen einlädt. Bei Schülern herrschen z.B. auch bei der Formel von Bayes erhebliche Verständnisschwierigkeiten. Zum Gelingen der oben vorgeschlagenen Änderungen ist es aber notwendig, dass vorher bereits der Satz von Bayes *richtig* verstanden wurde. Die oben vorgeschlagene Identifizierung als Inferenzmöglichkeit ist zwar für die spätere Gegenüberstellung notwendig, löst aber noch nicht die grundsätzlich bei dieser Formel bestehenden Verständnisschwierigkeiten. Wie man den Satz von Bayes auf leicht verständliche Art einführen kann, wird in *Anregungen zum Stochastikunterricht* (Borovcnik, Engel & Wickmann, 2001) erläutert. Unsere Arbeitsgruppe untersucht derzeit im Rahmen des DFG-Schwerpunktprogramms „Bildungsqualität von Schule", ob und wie sich im Stochastikunterricht Verständnisprobleme bei bedingten Wahrscheinlichkeiten in den Griff bekommen lassen. In einer Interventionsstudie am Gymnasium konzentrieren wir uns dabei auf die Frage, ob sich bedingte Wahrscheinlichkeiten im Allgemeinen und die Formel von Bayes im Speziellen mit Hilfe einer konsequenten Anwendung von absoluten Häufigkeiten verständnisfördernd unterrichten lassen. Anknüpfend daran soll eine Umsetzung des in diesem Artikel geschilderten Ansatzes zum Thema Signifikanztest in den Unterricht erfolgen. Wir hoffen, dass wir in einer der nächsten Ausgaben von „Stochastik in der Schule" von Ergebnissen dieses Projekts berichten können.

Literatur

Birnbaum, I. (1982). Die Interpretation statistischer Signifikanz (übersetzt von Günter Fillbrunn), Stochastik in der Schule, 2, 3.

Borovcnik, M., Engel, J. & Wickmann, D. (Hrsg.) Anregungen zum Stochastikunterricht: Die NCTM-Standards 2000, Klassische und Bayessche Sichtweise im Vergleich. Franzbecker: Hildesheim.

Buth, M. (1991). Zur Behinderung des gesunden Menschenverstandes durch Stochastik, Stochastik in der Schule, 3, S. 12-22.

Buth, M. (1993). Zum Thema Testen von Hypothesen: Was man aus der Forschungspraxis für die Schule lernen kann, Stochastik in der Schule, 2. S. 34-46.

Deschauer, S. (1999). Möglichkeiten einer historischen Akzentuierung im Mathematikunterricht, 15. Eichstätter Kolloquium zur Didaktik der Mathematik, S. 58/1-58/13.

Diepgen, R. (1985a). Was Schüler zum Hypothesentesten wissen sollten, Stochastik in der Schule, 1, S. 32-36.

Diepgen, R. (1985b). Signifikanz – Na und? Politik und Computer im Statistikunterricht. mathematik lehren, 13, S. 54-57.

Dinges, H. (1978). Schwierigkeiten mit der Bayesschen Regel. In: Mathematisch-

Physikalische Semesterberichte. Göttingen: Vandenhoeck & Ruprecht.

Falk, R., & Greenbaum, W. (1995). Significance tests die hard. Theory & Psychology, 5, 75-98.

Gigerenzer, G. (1993a). Über den mechanischen Umgang mit statistischen Methoden. In E. Roth (Hrsg.), Sozialwissenschaftliche Methoden. München: R. Oldenbourg Verlag.

Gigerenzer, G. (1993b). The Superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues. Hillsdale, NJ: Erlbaum.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1999). Das Reich des Zufalls. Spektrum.

Gigerenzer, G. & Krauss, S. (2001). Statistisches Denken oder statistische Rituale? Was sollte man unterrichten? In: Borovcnik, M., Engel, J. & Wickmann, D. (Hrsg.), Anregungen zum Stochastikunterricht: Die NCTM-Standards 2000, Klassische und Bayessche Sichtweise im Vergleich. Franzbecker: Hildesheim.

Haller, H. & Krauss, S. (eingereicht). Misinterpretations of Significance: A Problem Students share with their Teachers?

Krengel, U. (1990). Einführung in die Wahrscheinlichkeitstheorie und Statistik. Vieweg.

Lindley, D.V. (1965). Introduction to probability and statistics from a Bayesian viewpoint. Part 1: Probability, Cambridge University Press.

Nickerson, R.S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy, Psychological Methods, 5, 241-301.

Riemer, W. (1985). Neue Ideen zur Stochastik. Lehrbücher und Monographien zur Didaktik der Mathematik. Band 3. BI Wissenschaftsverlag.

Oakes, M. (1986). Statistical inference: A commentary for the social and behavioral sciences. Chichester: Wiley.

Stein, G. (1993). Schwierigkeiten mit der Nullhypothese. Stochastik in der Schule, 13, 1, S. 15-22.

Waldmüller, B. (1998). Was sagen signifikante Ergebnisse? – Zu einem Beispiel aus der Zeitung, Stochastik in der Schule, 18, 3, S.3-8.

Anschrift der Verfasser:
Stefan Krauss
Christoph Wassner
Max-Planck-Institut für Bildungsforschung
Lentzeallee 94
14195 Berlin
`{krauss;wassner}@mpib-berlin.mpg.de`