Statistische Tests in Gymnasiallehrbüchern

Dietrich Stoyan, Freiberg

Zusammenfassung: *Es wird diskutiert, wie Grundbegriffe der Theorie der Signifikanz-Tests in modernen deutschen Gymnasiallehrbüchern behandelt werden. Dabei wird auf Probleme hingewiesen, die aus Unterschieden in der Form der Formulierung der Nullhypothesen resultieren können. Schließlich wird bezweifelt, ob es sinnvoll ist, Alternativ-Tests im Gymnasium durchzunehmen.*

1 Einleitung

In vielen deutschen Bundesländern werden heutzutage in der Abiturstufe Signifikanztests behandelt. Der Verfasser hat seinerseits mehr als zwei Jahrzehnte lang Stochastik an einer Technischen Universität im Grundkurs für Ingenieurstudenten gelehrt. Da interessierte es ihn, wie dieser schwierige Stoff den Schülern vermittelt wird. Dazu inspizierte er vier einschlägige Lehrbücher, nämlich die der Verlage Cornelsen, Duden-Paetec, Klett (Lambacher Schweizer) und Westermann-Schroedel, alle für das Bundesland Sachsen. Im Folgenden wird über die Ergebnisse dieser Untersuchung kurz berichtet, die durchaus interessante Erkenntnisse erbrachte. Dabei werden auch kritische Bemerkungen gemacht, wobei aber die betroffenen Lehrbücher nicht konkret genannt werden. Wörtliche Zitate aus ihnen sind *kursiv* geschrieben. Die Stellen der Zitate in den Büchern können beim Autor erfragt werden.

Um für die folgende Diskussion eine sichere Grundlage zu haben, werden einige Sätze aus einem erfolgreichen Statistiklehrbuch für Studenten der Betriebwirtschaftslehre (Bamberg u. a., 2009, S. 159/160) zitiert, in denen Grundbegriffe zum Thema „Signifikanztests" vorkommen:

> „Über die Verteilung einer Grundgesamtheit liege eine Hypothese vor, in der etwa Erfahrungen, Vermutungen oder theoretische Überlegungen zum Ausdruck kommen, die aber auch nur in einer Behauptung bestehen kann, z. B. dass ein Verteilungsparameter einen vorgeschriebenen Sollwert einhält. Solche Hypothesen werden anhand der Ergebnisse einer Stichprobe überprüft. Dabei wird eine Hypothese als statistisch widerlegt angesehen und abgelehnt oder verworfen, wenn das Stichprobenergebnis in deutlichem (= signifikantem) Gegensatz zu ihr steht. Entsprechende Überprüfungsverfahren heißen daher Signifikanztests …"

Stochastik in der Schule 31 (2011) 1, S. 28–32

„Zu einer gegebenen Hypothese können verschiedene Alternativen bestehen. Wir bezeichnen die zu untersuchende Hypothese H_0 als die Nullhypothese und die jeweils relevante Alternative als Gegenhypothese oder Alternativhypothese H_1, …“

Signifikante Abweichungen des Verhaltens der Stichprobe von dem laut Nullhypothese erwarteten Verhalten sind auch dann möglich, wenn die Nullhypothese richtig ist. Dann würde man irrtümlich H_0 ablehnen.

„Wir fordern nunmehr, dass eine derartige Fehlentscheidung

Ablehnung von H_0, wenn H_0 richtig ist

lediglich mit einer als zulässig vorgegebenen kleinen Irrtumswahrscheinlichkeit α, dem so genannten Signifikanzniveau, vorkommen darf.“

Durch dieses α wird bestimmt, welche Abweichungen als signifikant anzusehen sind.

Vergleicht man die obige Erklärung des Signifikanztests mit der folgenden, in einem der Lehrbücher gegebenen, dann wird man mitleidig mit Schülern und Lehrern gestimmt:

In den meisten Fällen liegen die Dinge insofern etwas komplizierter, als dass man über den Wert von p nur eine Vermutung hat, die durch einen statistischen Test entweder bestätigt oder widerlegt werden soll. Einen solchen Test nennt man einen Signifikanztest.

Die Behandlung der Grundbegriffe

Nullhypothese, signifikant und Signifikanzniveau

in den verschiedenen Lehrbüchern ist der Inhalt des nächsten Abschnitts.

2 Grundbegriffe

Nullhypothese

In einem der Lehrbücher wird recht gut gesagt, dass *die Ausgangsvermutung zur Wahrscheinlichkeitsverteilung der Grundgesamtheit im Allgemeinen als Nullhypothese bezeichnet wird.*

In einem anderen heißt es aber: *Liegt eine Hypothese über eine Wahrscheinlichkeitsverteilung vor, so ist es sinnvoll, aufgrund eines Stichprobenergebnisses über eine mögliche Ablehnung dieser Hypothese zu entscheiden. Diese zu überprüfende Hypothese nennt man Nullhypothese, kurz H_0. – Ist H_0 wahr, so besteht kein („null“) Unterschied zwischen der angenommenen und der vorliegenden Verteilung.*

Die anderen Bücher schweigen sich über das Wort aus; es fällt einfach vom Himmel. Aber es wird auch nicht in Bamberg u. a. (2009) erklärt und auch nicht in des Verfassers eigenem Buch Stoyan (1992). In der Hochschullehre kann man offensichtlich Fragen der Studenten nach diesem Wort entgehen … Aber der Verf. kann sich nicht vorstellen, dass nicht alljährlich interessierte Schüler fragen, was dieses Wort soll.

So etwa könnte die Antwort auf die gedachte Schülerfrage lauten: Die Nullhypothese bildet den Ausgangs- oder Nullpunkt der statistischen Überlegungen, sie kennzeichnet den Fall von Null-Interesse, den Fall von „schon bekannt“, den Fall von Null-Abweichung von den bisher unterstellten Annahmen …

Ideal zur konkreten Erläuterung ist die in den Lehrbüchern behandelte Nullhypothese $H_0: p = 0{,}5$ für die Wappenwahrscheinlichkeit beim Münzwurf. Diese Hypothese ist es natürlich, mit der man Untersuchungen über die Münzwurfwahrscheinlichkeiten beginnt, und es würde erst dann interessant, wenn man Abweichungen nachweisen könnte.

Signifikant

Noch einmal zur Erinnerung: Signifikant können Abweichungen sein. Man sagt etwa:

Die Abweichung vom Sollwert ist signifikant, die nachgewiesene Korrelation ist signifikant, oder die Wappenwahrscheinlichkeit weicht signifikant von 0,5 ab.

Dagegen kann ein Test nicht die Eigenschaft haben, signifikant zu sein, und ein mit $\alpha = 0{,}01$ durchgeführter ist auch nicht „hochsignifikant“, wie in einem der Lehrbücher geschrieben. Das mag als nachlässiger Statistiker-Jargon durchgehen, aber in einem Gymnasiallehrbuch haben solche Formulierungen keinen Platz. (Auch Seife kann nicht signifikant sein, aber die Verwendung von Seife beim Händewaschen kann einen signifikanten Effekt haben.)

Erstaunlich ist, dass das Wort „signifikant“ außer in dieser fehlerhaften Art und Weise (in einem Lehrbuch) nirgendwo so ausführlich erklärt wird, wie es sinnvoll wäre. Dabei zeigen die Verfasser der Lehrbücher doch, dass sie das Wort richtig gebrauchen können. Zum Beispiel liest man:

… signifikante Abweichungen könnten dabei nach unten oder nach oben auftreten

und

Wenn das Stichprobenergebnis signifikant von $p = 0{,}11$ *und von jedem* $p < 0{,}11$ *abweicht, dann verwerfen wir …*

Signifikanzniveau

Das Signifikanzniveau wird in allen Lehrbüchern korrekt erklärt als die obere Schranke für die Wahrscheinlichkeit des Fehlers 1. Art. Das war zu erwarten, weil sich alle Bücher auf die Berechnung der Wahrscheinlichkeiten der Fehler beim Testen konzentrieren. Allerdings wird der Name nirgendwo erklärt, weil eine entsprechende Erklärung auf das Wort „signifikant“ Bezug nehmen müsste.

3 Zwei Formen der Nullhypothesen

In den Lehrbüchern wie auch in der statistischen Fachliteratur gibt es zwei Versionen, die Nullhypothesen zu formulieren. Das stört in der Anwendung der Tests nicht weiter, kann aber im Schulbetrieb zu schweren Problemen führen, nämlich dann, wenn Aufgaben in zentralen Prüfungsarbeiten im Sinne einer der beiden Versionen gestellt werden.

Aus Bamberg u. a. (2009), S. 159, wird das folgende Beispiel leicht modifiziert übernommen.

Eine Fabrik erzeugt Alkopop-Getränke mit einem Alkoholgehalt von 5 %. Dabei treten beim Alkoholgehalt in den Flaschen leichte Schwankungen auf. Die Hypothese H_0, dass der Alkoholgehalt gleich dem Sollwert von $p_0 = 0{,}05$ ist, soll anhand einer Stichprobe überprüft werden. Auf Grund der Interessenlage derjenigen Personen, die die Untersuchung vornehmen, sind drei Fälle zu unterscheiden, nämlich: Die Überprüfung geschieht durch

a) eine Eichkommission, die an einer Abweichung vom Sollwert $p = 0{,}05$ 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 zuviel 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.

Es gibt nun zwei Versionen, diese Test-Situationen in Null- und Alternativhypothese H_0 und H_1 auszudrücken.

Version 1

$H_0 : p = p_0$

und

a) $H_1 : p \neq p_0$

b) $H_1 : p < p_0$

c) $H_1 : p > p_0$.

Version 2

a) $H_0 : p = p_0, H_1 : p \neq p_0$

b) $H_0 : p \geq p_0, H_1 : p < p_0$

c) $H_0 : p \leq p_0, H_1 : p > p_0$.

Der Verfasser hat in seinen Lehrveranstaltungen bei der Einführung der Signifikanztests die Version 1 bevorzugt. Bamberg u. a. (2009) folgen ebenfalls diesem Weg, während stärker mathematisch-theoretische Bücher fast durchweg die Version 2 behandeln. Dabei merken Bamberg u. a. (2009) beiläufig auf S. 164 an, dass der Übergang zu Version 2 leicht möglich ist. Von den vier inspizierten Lehrbüchern folgen zwei der Version 1 (Lambacher Schweizer und Cornelsen) und zwei der Version 2 (Duden-Paetec und Schroedel).

Eines der Bücher, das der Version 1 anhängt, rechtfertigt seine Vorgehensweise recht geschickt wie folgt: *Wenn nur bekannt ist, dass die Trefferwahrscheinlichkeit bisher mindestens* $p = 5\ \%$ *betrug, so lautet die Nullhypothese* $H_0 : p \geq 0{,}05$. *Es genügt hier jedoch, den ‚Extremfall'* $p_0 = 0{,}05$ *zu betrachten. Je weiter nämlich die Trefferwahrscheinlichkeit über 5 % liegt, desto unwahrscheinlicher ist es, sich fälschlicherweise für die Gegenhypothese* $H_1 : p < 0{,}05$ *zu entscheiden.*

Es wird nun gezeigt, welche Konsequenzen die beiden Versionen in Prüfungsarbeiten haben können.

In der Abiturprüfung 2007 des Landes Nordrhein-Westfalen wurde für den Leistungskurs Mathematik folgende Aufgabe gestellt:

Die Firma Reinlich und Sohn möchte den Bekanntheitsgrad p ihres Waschmittels Reinil ermitteln. Eine erste Befragung von 50 Personen deutet auf einen Bekanntheitsgrad von $p = 0{,}4$ hin. Die Firma erwägt, für ihr Produkt im Fernsehen zu werben. Wegen der hohen Kosten soll mittels einer Befragung von 800 Personen die Notwendigkeit der Werbemaßnahme überprüft werden. Herr Reinlich Senior und Herr Reinlich Junior sind sich einig, dass die Werbemaßnahme nur durchgeführt werden soll, wenn der Bekanntheitsgrad des Waschmittels unter 40 % liegt. Keine Einigkeit erreichen die beiden bei der Wahl der Nullhypothese.

a) Herr Reinlich Junior möchte die Hypothese $H_0 : p \leq 0{,}4$ überprüfen

…

b) Herr Reinlich Senior möchte lieber die Hypothese $H_0 : p \geq 0{,}4$ überprüfen.

Diese Aufgabenstellung ist ganz im Sinne von Version 2 formuliert. Schüler, die im Sinne von Version 1 unterrichtet wurden, werden kaum die Aufgabenstellung verstehen. Im Sinne von Version 1 würde die Aufgabe etwa folgendermaßen lauten:

… wenn der Bekanntheitsgrad des Waschmittels unter 40 % liegt. Während sich beide über die Nullhypothese $H_0 : p = 0{,}4$ einig sind, haben sie über die Alternativ-Hypothese verschiedene Ansichten.

a) Herr Reinlich Junior bevorzugt die Alternative $H_1 : p > 0{,}4$

…

b) Herr Reinlich Senior möchte lieber die Alternative $H_1 : p < 0{,}4$ überprüfen.

Hier kämen nun die Schüler, die nach Version 2 unterrichtet wurden, in Schwierigkeiten.

Die Konsequenz für ein Bundesland mit Zentralabitur, in dem beide Versionen unterrichtet werden, kann wohl nur lauten, dass im Abitur Aufgaben mit einseitigen Tests vermieden werden sollten oder dass zwei Formulierungen derselben Aufgabe angeboten werden müssen.

Oder ist es denkbar, dass sich die deutschen Stochastik-Lehrer auf eine der beiden Versionen einigen?

4 Alternativ-Test

Abschließend wird noch kurz der Alternativ-Test diskutiert, der in Sachsen in den Gymnasien behandelt wird. Wie der Verfasser vernahm, soll das didaktisch sinnvoll sein, weil bei diesem Testtyp die verschiedenen Fehler (1. und 2. Art) besonders klar berechnet werden können.

Eine typische Aufgabe, die mit dem Alternativ-Test behandelt wird, lautet folgendermaßen:

Eine Agrargenossenschaft erhält mehrere Container mit Saatgetreide. Bestellt waren zwei Sorten: Während die eine Sorte eine Keimfähigkeit von 90 % besitzt, dafür aber anfälliger gegenüber Schädlingen ist, keimt die andere nur in 75 % der Fälle, ist aber gegenüber einem Schädlingsbefall resistenter.

Aus der Beschriftung der angelieferten Container ist allerdings nicht mehr zu ersehen, welches Saatgetreide sich in welchem Container befindet. Es soll der Inhalt eines bestimmten Containers mithilfe einer Stichprobe von 100 Getreidekörnern untersucht werden, indem man diese Körner in einem Gewächshaus zum Keimen bringt.

Es ist klar, dass die Anzahl k der gekeimten Körner die entscheidende Größe ist; ist sie groß, wird man annehmen, dass die Keimfähigkeit 90 % beträgt, andernfalls wird man annehmen, dass sie gleich 75 % ist.

Hätte der Autor diese Aufgabe seinen Studenten gestellt, hätte er noch den erreichten Wert von k angegeben und gefragt, welche konkrete Schlussfolgerung die Agrargenossenschaft zieht. Es soll hier angenommen werden, dass $k = 82$ ist.

Alle vier Lehrbücher gehen solche Aufgaben auf einem höheren mathematischen Niveau an und konstruieren einen Test, der die Hypothesen

$$H_a : p = p_a \quad \text{und} \quad H_b : p = p_b$$

überprüft und lösen die Aufgabe für allgemeines k. Dabei ist die Verfahrensweise so, dass die Ablehnung von H_a automatisch die Annahme von H_b bedeutet.

Übrigens wurde hier die Indizierung „a“ und „b“ gewählt, um die üblichen Indizes „0“ und „1“ zu vermeiden, da ja keinerlei Bezug zu einer Nullhypothese im Sinne der Signifikanztests besteht. (Was sollte denn bei dem obigen Beispiel die Nullhypothese sein?) Dabei ist zu beachten, dass bei Aufgaben des obigen Typs die Indizes „a“ und „b“ ganz willkürlich vergeben werden können, das Endergebnis des Tests ist immer dasselbe. Das wird in den meisten Lehrbüchern nicht so explizit gesagt. In einem werden aber die Indizes „1“ und „2“ gewählt.

Der Alternativ-Test wird in der Statistik-Lehre an den Hochschulen nur selten behandelt, sicher weil die Signifikanztests als viel wichtiger angesehen werden und weil Alternativ-Tests nur selten angewendet werden. Für Mathematik-Studenten hat das dahinter stehende Neyman-Pearson-Lemma allerdings grundlegende Bedeutung.

Ein normaler angewandter Statistiker würde die obige Aufgabe mit den Getreidekörnern ganz anders und viel eleganter lösen. Er würde die Aufgabe als Klassifizierungsproblem (oder als Problem der Diskriminanzanalyse) auffassen, wo es darum geht, die Stichprobe der 100 Körner einer der Klassen „75“ und „90“ zuzuordnen. Dazu wird die Likelihood $L_k(p)$ für $k = 82$ gekeimte Samen bei 100 gesäten Körnern bei der Keimfähigkeit von p berechnet. Die „Likelihood“ ist die Wahrscheinlichkeit „für das Beobachtete“.

Im gegebenen Fall gilt

$$L_k(p) = p^k \cdot (1-p)^{(n-k)}$$

mit $n = 100$. k-mal keimt ein Korn, $(n-k)$-mal nicht. (Es ist kein Versehen, dass hier der vielleicht erwartete Binomialkoeffizient $\binom{n}{k}$ fehlt.)

Nun werden mit $k = 82$ und $n = 100$ die Likelihood-Werte $L_{82}(0{,}75)$ und $L_{82}(0{,}9)$ berechnet. Dem p-Wert mit der größeren Likelihood wird die Stichprobe zugeordnet.

Es gilt

$$L_{82}(0{,}75) = 8{,}27 \cdot 10^{-22}$$

und

$$L_{82}(0{,}90) = 1{,}77 \cdot 10^{-22}.$$

Also schlussfolgert man, dass die Probe zur Klasse „75“ gehört.

Um ein Missverständnis zu vermeiden: Der Autor möchte nicht dafür plädieren, die Idee der Likelihood im Schulunterricht zu behandeln.

Literatur

Bamberg, G., Baur, F., Krapp, M. (2009): Statistik. München: Oldenbourg Verlag.

Bigalke, A., Köhler, V. (Hsg.) (2007): Mathematik, Analytische Geometrie – Stochastik, Band 2, Berlin: Cornelsen Verlag.

Bossek, H., Heinrich, R. (Hsg.) (2008): Mathematik, Jahrgangstufe 12, Sachsen. Mannheim: Duden-Paetec Verlag.

Griesel, H., Gundlach. A., Postel, H., Suhr, F. (Hsg.) (2009): Elemente der Mathematik. Sachsen, 12. Schuljahr Grund- und Leistungskurs, Braunschweig: Bildungshaus Schulbuchverlage, Westermann Schroedel Diesterweg Schöningh Winklers GmbH.

Lind, D., Negwer, J., Neumann, P. (Hsg.) (2008): Lambacher Schweizer, Gesamtband 11/12, Mathematik für Gymnasien Mathematisches Unterrichtswerk für das Gymnasium, Ausgabe Sachsen, Stuttgart, Leipzig: Ernst Klett Verlag.

Stoyan, D. (1993): Stochastik für Ingenieure und Naturwissenschaftler. Berlin: Akademie Verlag.

Anschrift des Verfassers

Prof. Dr. Dietrich Stoyan
TU Bergakademie Freiberg
Institut für Stochastik
Prüferstr. 9
09599 Freiberg
stoyan@math.tu-freiberg.de