Wieviel ausgezeichnete Schulnoten sind tolerierbar?

VON HENRIK DAHL, ÜBERSETZT UND BEARBEITET VON GERHARD KÖNIG

Originaltitel: How Many Excellent Grades Should be Tolerated

Teaching Statistics, volume 23(Spring 2001)1, S. 24-25

Zusammenfassung: *Der Artikel modelliert eine interessante politische Schulsituation, die die Ideen des Testen von Hypothesen verdeutlichen kann.*

Einleitung

Das Testen statistischer Hypothesen ist ein schwieriger Unterrichtsgegenstand. Für den Lehrer ist es bedeutsam, den Schülern und Studenten sowohl die Möglichkeiten als auch die Grenzen dieser Methode aufzuzeigen. Daher ist es wichtig, Beispiele anzubieten, die einerseits interessant erscheinen, andererseits charakteristisch sind. Das folgende Beispiel genügt hoffentlich beiden Voraussetzungen.

Das Beispiel

Die Schulnote "ausgezeichnet" oder "sehr gut" wird in norwegischen Schulen nur Schülern des oberen 4%-Bereichs vergeben. Dabei geht man von einer großen Menge (mehrere Tausend) von Schülern aus.

Wieviel "ausgezeichnet" sollen in einer Schulklasse mit 25 Schülern akzeptiert werden ohne dass der Lehrer als "Softie" oder "Weichei" bezeichnet wird? (sowohl die Klassengröße als auch die Softie-Grenze können natürlich modifiziert werden.)

In der Sprache der Testtheorie heißt das, die Nullhypothese $p = 0.04$ wird gegen die Alternativhypothese $p>0.04$ getestet, wobei p die Wahrscheinlichkeit ist, dass ein (zufällig ausgewählter) Schüler dieser Klasse ein "ausgezeichnet" bekommt. Die Teststatistik X definiert die Anzahl der "ausgezeichnet" in einer Klasse mit 25 Schüler. Die Nullhypothese zu verwerfen (d.h. den betr. Lehrer als zu sanft lächerlich zu machen) im Fall $X\geq4$, gibt ein Signifikanzniveau von 1.65%. Dies wurde mittels der Binomialverteilung ermittelt. Die Modellannahme der gleichen Wahrscheinlichkeit für "ausgezeichnet" der 25 Schüler und ihrer Unabhängigkeit bei den 25 Schülern mag nicht realistisch erscheinen, dennoch kann sie als erste Modellannahme bestehen. Da sehr wenig Lehrer es wagen würden - in Norwegen- mehr als vielleicht zwei "ausgezeichnet" in einer Klasse mit 25 Schülern zu vergeben, zeigt dies, dass Lehrer vielleicht überreagieren und denken, dass das Gesetz der großen Zahlen auch auf kleine Zahlen, wie z.B. 25 anwendbar sei (Kahnemann et al.1982). Es kann argumentiert werden, dass ein realistischeres Modell den kritischen Wert des Tests wegen der Heterogenität der Klassen und möglicher Abhängigkeiten erhöhen würde.

Dieses Beispiel demonstriert einige Probleme beim Hypothesentesten. Wenn das Signifikanzniveau bei 1.65% festgelegt wird, ist der Test schwach. Für $p = 0.1$ ist können wir die Binomialverteilung benutzen, um auszurechnen, dass die Macht des Tests nur 23.64% ist, was eine Wahrscheinlichkeit von 76.36% für einen Fehler zweiter Art bedeutet. $P=0.1$ bedeutet, dass der Lehrer 10% der Schüler ein "ausgezeichnet" gibt, was sein Ansehen bei den Schülern heben könnte aber nicht dem allgemeinen Standard genügte.

Die geringe Macht des Tests könnte ein Argument dafür sein, den kritischen Wert zu erniedrigen. Machen wir es: "Lächerlich weich wenn $X\geq3$". Das Signifikanzniveau dieses Tests wäre 7.65%. Dies würde bedeuten, dass 7.65% der Lehrer, die sich korrekt an die Standards hielten, als Softies oder Weicheier bezeichnet würden. Es ist klar, dass ein solcher Test nicht akzeptabel ist. Mit nur 25 Schülern in einer Klasse sind solche Überlegungen also nicht durchführbar.

Betrachten wir nun einen Lehrer, der 10 Klassen mit 25 Schüler zu benoten hat, also insgesamt 250 Schüler. Wie oben gehen wir vereinfachend davon aus, dass die Zufallsvariable Y, definiert als die Anzahl der "ausgezeichnet" unter den 250 Schülern, binomialverteilt ist mit den Parametern $n=250$ und $p=0.04$, unter der Nullhypothese einer fairen Notengebung. Der dazugehörige Test, $p=0.04$ zurückzuweisen wenn $Y\geq18$, hat ein Signifikanzniveau von ungefähr 0.8%.

Ist es vernünftig, diese Irrtumswahrscheinlichkeit in dieser Situation zu akzeptieren? Bevor wir darüber entscheiden können, müssen die Irrtumswahrscheinlichkeiten der Fehler zweiter Art betrachtet werden. Dies bedeutet, Lehrer in Betracht zu ziehen, die die Standards durchbrechen. Die Macht des Tests "Softie, wenn $Y \geq 18$" bei $p=0.01$ ist etwa 94%, was bedeutet, dass die Wahrscheinlichkeit eines Fehlers zweiter Art etwa 6% ist. Einen Lehrer ungerechtfertigterweise als zu sanft beim Notengeben zu verdächtigen, erscheint weniger schwerwiegend als einen Lehrer mit $p=0.1$ akzeptieren zu müssen. Daher erscheint die Entscheidungsregel "Weichei, wenn $Y \geq 18$" doch vernünftig.

Das dargestellte Beispiel sollte Schülern und Studenten helfen, die Ideen und Konsequenzen von Fehlern erster und zweiter Art zu verstehen, zu sehen, dass die Wahrscheinlichkeiten dieser beiden Fehler zusammenhängen und die Bedeutung der Stichprobengröße zu erkennen.

Literatur

Batanero, C. (1997):Should we get rid of statistical testing? The significance test controversy. In: ISI Newsletter v.21(2), S. 19

Buth, M.: Zum Thema 'Testen von Hypothesen': was man aus der Forschungspraxis für die Schule lernen kann. In: Stochastik in der Schule (1993) v. 13(2) S. 35-46

Bruhn, J.: Retrospektiver Einsatz von klassischen Testverfahren. Eine erkenntnistheoretische Fragestellung beim Testen von Hypothesen. In: Stochastik in der Schule. (1986) v. 6(2) S. 5-12

Danckwerts, R.; Vogel, D.: Das Testen von Hypothesen - Missverständnisse und Perspektiven. In: Didaktik der Mathematik (1993) v. 21(1) S. 51-65

Danckwerts, R.; Vogel, D.: Das Testen von Hypothesen - Missverständnisse und Perspektiven. In: Didaktik der Mathematik (1993) v. 21(1) S. 51-65

Hofmann, W. : Das Testen von Hypothesen. München: Bayerischer Schulbuch-Verlag. 1986. 191 S.

Kahnemann, D.; Slovic, P.; Tverski, A. (1997): Judgment under Uncertainty: Heuristics and Biases. New York: Cambridge University Press

Ulshöfer, Klaus: Zum Testen von Hypothesen. In: Mathematik in der Schule. (Dez 1997) v. 35(12) S. 663-670, 675-680

Unterrichtsideen zum Thema Stochastik

Ein "Sonderforschungsbereich" Stochastik ist im Arbeitsbereich des BLK-Modellversuchs "SelMa" eingerichtet worden. Das neue Angebot versteht sich als Begegnungsplattform für Lehrerinnen und Lehrer, die sich für dieses Gebiet der Mathematik interessieren. In der "gläsernen Werkstatt" finden sie anwendungsorientierte Anregungen und Ideen für die Unterrichtsgestaltung. Gleichzeitig rufen die Betreuerinnen und Betreuer des Arbeitsbereich Besucher dazu auf, eigene Unterrichtseinheiten aus dem Bereich der Stochastik im "Sonderforschungsbereich" zu veröffentlichen. Dabei ist es zweitrangig, ob die eigenen Ideen ganz zu Ende gedacht sind. Auch unfertige Anregungen und Vorschläge sind willkommen. Ansprechpartner für Tipps und Nachfragen zum Thema ist Rainer Altmann.

"SelMa" ist die Abkürzung für den BLK-Modellversuch "Selbstlernen in der gymnasialen Oberstufe - Mathematik", der vom Landesinstitut für Schule und Weiterbildung in Soest durchgeführt wird. Autorenteams an fünf nordrhein-westfälischen Schulen entwickeln Unterrichtseinheiten für den Einsatz Neuer Medien im Mathematikunterricht in der gymnasialen Oberstufe. An zehn so genannten Erproberschulen werden die neuen Unterrichtsszenarien auf ihre Praxistauglichkeit hin getestet.

http://www.learn-line.nrw.de/selma/medio/stochastik