# Eine Einführung in das Testen von Hypothesen[1]

BART K. HOLLAND, NEW JERSEY, USA

[1] Original: A Classroom Demonstration of Hypothesis Testing.
In *Teaching Statistics* 29 (2007) 3, 71–73.
Kürzung, Bearbeitung und Übersetzung: JÖRG MEYER

***Zusammenfassung:*** *Es geht um eine Vorführung, die das Verständnis und die Motivgebung in Bezug auf das Hypothesentesten verbessert.*

## 1 Hintergrund und Ziele

Viele Lernende empfinden bei der ersten Begegnung Hypothesentests als abstrakt und kontraintuitiv. Es muss das Ziel der Lehrperson sei, alle mit Hypothesentests zusammenhängende Schwierigkeiten zu überwinden, so dass die Lernenden im weiteren Verlauf t-Tests und Chi-Quadrat-Tests mit Verständnis und nicht nur als mechanischen Prozess anwenden können. Zwar könnte das reine Auswendiglernen einer Testprozedur es Lernenden ermöglichen, den Test durchzuführen und auch das Ergebnis mit dem passenden Statistik-Jargon papageienhaft nachzuplappern, aber Lehrende sollten sich schon darum kümmern, dass die Prozeduren auch wirklich verstanden werden.

Eine gute praktische Vorführung kann bei einer abstrakten Behandlung des Hypothesentestens die hinter dem Prozess steckende Logik entmystifizieren und auch den Lernenden einige Ängste nehmen. Die Vorführung kann auch dazu diesen, dass die Lernenden die Begrifflichkeit und die Terminologie besser beherrschen, weil sie Begriffe mit erlebten Inhalten verbinden.

Der vorliegende Artikel beschreibt eine Vorführung zum Hypothesentests, die ich im Laufe der Jahre immer weiter verfeinert habe. Seitdem beklagt sich keiner mehr bei mir, das Hypothesentesten nicht verstanden zu haben.

## 2 Die Vorführung

Man benötigt eine Münze mit zwei gleichen Seiten (also zwei Köpfen) für die Vorführung; man kann solche Münzen über das Internet erwerben.

Die Vorführung beginnt damit, dass man den Lehrenden gegenüber deutlich macht, dass sie immer schon Hypothesentests angewendet haben und dass nichts Esoterisches oder Ungewohntes dabei sei, auch wenn die Begriffe manchmal diesen Eindruck machen sollten. Sie würden jetzt gleich eine Hypothese statistisch testen, und zwar ohne jegliche Vorbereitung.

Man braucht einen Freiwilligen unter den Lernenden, der den übrigen mitteilt, ob ein Münzwurf „Kopf" oder „Zahl" zeigt. Man sage, dass es gleich eine Reihe von Münzwürfen gebe und dass daher eine neutrale Person (der Freiwillige) den anderen die Ergebnisse mitteilen solle, um nicht als Versuchsdurchführer in Verdacht zu geraten, falsche Angaben zu machen. Man hole eine Handvoll Münzen aus der Tasche, worunter sich die Münze mit den beiden gleichen Seiten befindet. Man wähle diese gezinkte Münze aus, so dass die Lernenden den Eindruck haben, diese Münze sei ungezinkt und sei zufällig unter mehreren ausgewählt. Nun können die Münzwürfe beginnen.

Man werfe die Münze so, dass sie auf der eigenen Hand zu liegen kommt. Der Freiwillige besieht sich jeweils das Ergebnis (darf die Münze aber nicht anfassen oder gar umdrehen) und schreibt es an die Tafel. Das mache man, bis man 10-mal „K" erreicht hat, ohne dass man Überraschung oder Neugier zum Ausdruck bringt. Während der Durchführung werden die Lernenden Bemerkungen wie „keine faire Münze" oder „Münze hat zwei gleiche Seiten" machen, auf die man als Lehrperson nicht reagieren darf. Nach den 10 Würfen stecke man die Münze wieder in seine Tasche.

## 3 Erklärungen

Man sage den Lernenden, dass die Vorführung zu allen Elementen eines statistischen Tests einer Hypothese geführt habe. Man frage die Lernenden, ob sie irgendwelche Schlussfolgerungen darüber hätten, ob die Münze fair war oder nicht, und lasse sie diese erläutern. Mit großer Sicherheit wird man die Antwort bekommen, dass 10-mal „Kopf" nacheinander doch sehr unwahrscheinlich sei, und dass die Wahrscheinlichkeit nur so groß sei wie $(1/2)^{10} = 0{,}000977$. Man würde also nur in etwa jeder 1000. Wurfserie ein solches Ergebnis sehen.

Dann kann man fragen, woher der Wert 1/2 in der Klammer oben kommt. Die Lernenden kommen darauf, dass dieser Wert von der Annahme herrührt, es würde sich um ein faire Münze handeln. Nun kann man die Begriffe „p-Wert", „Nullhypothese" und „Alternativhypothese" an die Tafel schreiben. Hypothesentests arbeiten mit dem p-Wert, also der Wahrscheinlichkeit, die Daten, die man gesehen hat, zu erhalten. Aber die Rechnung musste von irgendeiner Annahme ausgehen, und die Lernenden haben automatisch die Annahme der Nullhypothese gewählt. Da der erhaltene p-Wert sehr klein ist, ist es sehr unwahrscheinlich, die gesehene Sequenz von „K"s unter der Annahme der Nullhypothese erhalten zu haben.

Es wird nun für die Lehrperson gut sein, eine formale korrektere und vollständige Definition des p-Werts einzuführen: Der p-Wert umfasst die Wahrscheinlichkeit für die beobachtete Sequenz und für (unter Annahme der Nullhypothese) noch extremere. Dies führt ganz natürlich zu den „Schwanzbereichen" der Verteilung als den für die Entscheidung kritischen Bereichen. Die Wahrscheinlichkeit, extreme oder noch extremere Ausgänge zu beobachten, hängt natürlich davon ab, ob der Test einseitig oder zweiseitig ist.

Man mache deutlich, dass verschiedene Lernende verschiedene Ansichten darüber haben, ab wann sie davon überzeugt sind, es mit einer gezinkten Münze zu tun zu haben, also verschiedene alpha-Niveaus haben. Wenn der p-Wert kleiner ist als dieses alpha, wird die Nullhypothese zurückgewiesen. Man kann hier deutlich machen, dass alpha recht willkürlich ist, wobei Konventionen zu gewissen alpha-Werten geführt haben: Man sieht meistens Werte von 5 %, 1 % und 1 ‰ in wissenschaftlichen Texten. Es wird als fair betrachtet, wenn man einen Plan vorab eines Forschungsprojekts aufstellt, so dass alle betroffenen Parteien sich vorab darauf einigen, was als „Beweis" für einen Sachverhalt gelten soll; dies bedeutet, sich auf einen alpha-Wert im Voraus zu einigen und diesen auch zu veröffentlichen, damit andere Personen den Grad der Evidenz beurteilen können.

Mitunter erscheint die Natur des Hypothesentests als pervers gegenüber Nichteingeweihten: Das Verfahren ist zentriert darauf, etwas von geringem Interesse zurückzuweisen, anstatt die Alternativhypothese, die i. a. von viel größerem Interesse ist, in den Vordergrund zu stellen. Die geschilderte Vorführung zeigt jedoch, dass der Hypothesentest so vorgeht wie jedermann in seinem Alltagsschließen. In folgenden Abschnitten erkläre ich, dass wir in der Vorführung gewisse Wahrscheinlichkeiten für verschiedene Datensätze ausgerechnet haben. Dies führt zur Binomialverteilung, zur Normalverteilung, zur t-Verteilung und vielen anderen.

## Anschrift des Verfassers

Bart K. Holland
New Jersey Medical School, USA
holland@njms.rutgers.edu