

Hypothesen – schnell bei der Hand!¹

LAWRENCE M. LESSER, EL PASO, TX, USA

¹ Original ‚High-speed hypotheses!‘ in *Teaching Statistics* 34 (2012) 1, 10–12.

Übersetzung und Bearbeitung:

MANFRED BOROVCNIK, KLAGENFURT

Zusammenfassung: Ein interaktiver Zugang zum Hypothesentest wird vorgestellt. Weil die Studierenden in die Arbeit miteinbezogen werden, gestaltet sich Lernen sehr effizient. Schlagworte: Unterrichten von Hypothesentesten, p -Wert.

1 Einleitung

Wo doch die Unterrichtszeit so wertvoll ist, hilft es sehr, ein Repertoire von erprobten und anregenden Aktivitäten bei der Hand zu haben, die allesamt in fünf Minuten durchführbar sind – von der Beschaffung der Daten bis zur Ausrechnung des p -Werts. Solche Experimente stützen die Empfehlungen von Arbeitsgruppen der American Statistical Association, „reale Daten zu nützen“ und „aktives Lernen zu ermöglichen“ (Franklin et al. 2007), den statistischen Problemlöseprozess zu illustrieren (Aliaga 2010) und das Potential auszunutzen, das darin liegt, Spaß am Tun in die Klasse zu bringen (Lesser & Pearl 2008). In den folgenden Experimenten wird eine repräsentative Bandbreite an Aktivitäten zum elementaren Hypothesentest abgedeckt.

2 Aktivität 1: Wer hat ESP?

Hintergrund: Außersinnliche Wahrnehmung (extra-sensory perception – ESP) wird breit diskutiert und die Studierenden finden ein Interesse daran zu klären, ob sie selbst eine solche besondere Wahrnehmung haben. Einige Lehrbücher wie Utts (2005) verwenden ESP-Experimente in Fallstudien. Solche Experimente basieren im Kern auf Vorhersagen in einer kleineren Anzahl von Items (vier oder fünf Optionen), aber das kann für unterrichtliche Zwecke – wie hier – auch modifiziert werden.

Statistisches Thema: (Einseitige) Hypothesen über einen Anteil p .

Ablauf: Ein vollständiger Satz von 52 Karten wird gebraucht. Man vergewissere sich, dass die Studierenden mit der Aufteilung der Farben und der Kartenwerte (Benennungen) vertraut sind. Die Karten werden mit Zurücklegen gezogen (und dazwischen gemischt). Die Studierenden notieren ihre Vorhersa-

gen für z. B. ein Dutzend Karten. Unter der Nullhypothese „kein ESP“ beträgt die Wahrscheinlichkeit $1/52 \approx 1,9\%$, eine einzelne Karte richtig vorherzusagen. Während nun die Karten nacheinander gezogen werden, konzentrieren sie sich und machen unabhängig voneinander ihre Vorhersagen.

Die meisten von ihnen werden nicht mehr als zwei richtige Vorhersagen haben. Daher sind die p -Werte rasch auszurechnen. Wenn man vorweg fragt, ob jemand über ESP verfügt, so wird das wohl kaum jemand in der Klasse mit „ja“ beantworten. Daher wird man in einer Klasse, wenn sie nur groß genug ist, wenigstens einen erwarten, der einen persönlichen p -Wert von weniger als 5 % erreicht, was in diesem Fall zur fälschlichen Verwerfung der (zutreffenden) Nullhypothese führt.

Als Lehrender könnte man auch versucht sein, einen „magischen Trick“ einzubringen (etwa die Karten unauffällig markieren) und damit fast alle oder wenigstens sehr viele Karten (damit der Trick nicht zu offensichtlich wird) erraten und so einen ganz kleinen p -Wert bekommen. Lesser & Glickman (2009) diskutieren den Wert solcher Tricks im Statistikkunterricht. Will man die Erfolgswahrscheinlichkeiten variieren, so kann man die Vorhersagen auf die Farbe oder den Kartenwert beziehen. Auf der Internet-Seite von Boundary Institute (o. D.) können die Studierenden weiter forschen und testen, ob sie über außersinnliche Wahrnehmung verfügen oder nicht. Es geht hierbei um die Vorhersage von Karten, Bildern, Orten, Lotteriezahlen etc.

3 Aktivität 2: Nimm einen Sitzplatz ein

Hintergrund: Es gibt einige Literatur über Präferenzen, etwa wo sich Personen in Abhängigkeit von der Art des Raumes bevorzugt hinsetzen (z. B. Kaya & Burgess 2007); Geschlecht wird hier auch als potentielle Einflussgröße genannt. Für dieses Experiment nehmen wir an, dass die Bestuhlung der Klasse in Form eines Rechtecks vorgegeben ist und die Plätze durch einzelne Sessel in Reihen angeordnet sind.

Statistisches Thema: Test der Nullhypothese, ob zwei Anteile gleich sind (auch in der Form des Chi-Quadrat-Tests auf Unabhängigkeit).

Ablauf: Die Studierenden haben – bei freier Platzwahl – beim Eintreten ins Klassenzimmer ihren Platz eingenommen. Jetzt sollen sie durch Diskussion zu einer

operationalen Definition von „vordere“ und „hintere Hälfte“ des Raums kommen. Sind diese Begriffe geklärt, so können sie in weniger als einer Minute die Daten ihrer eben eingenommenen Sitzordnung in die Tabelle 1 eintragen und testen, ob Männer und Frauen dieselbe Wahrscheinlichkeit haben, einen Platz im vorderen Teil der Klasse einzunehmen oder nicht.

Dazu sind auch die beiden Merkmale der Population zu beachten und es ist hervorzuheben, dass die gegenwärtige Sitzordnung dahingehend untersucht wird, ob sie als zufällige Stichprobe betrachtet werden kann.

| | Anzahl der Männer | Anzahl der Frauen |
|----------------|-------------------|-------------------|
| Vordere Hälfte | | |
| Hintere Hälfte | | |

Tab. 1: Arbeitsblatt zum Festhalten der gewählten Sitzordnung in der Klasse

4 Aktivität 3: Einen Ton halten

Hintergrund: Einige Forschungsarbeiten zeigen, dass man stehend besser als in sitzender Position singen kann (etwa Cheng 1991), andere Arbeiten verweisen wiederum darauf, dass es weit mehr darauf ankommt, wie gut man sich positioniert, unabhängig davon, ob man steht oder sitzt.

Statistisches Thema: Test von zwei Mittelwerten auf Gleichheit.

Ablauf: Durch einen Münzwurf wird entschieden, ob man der Sitzgruppe (Zahl) oder der Stehgruppe (Kopf) angehört. Eine Stoppuhr wird auf eine Leinwand projiziert, sodass sie von allen gut eingesehen werden kann (etwa online *Stopwatch* im Menü *Interactivate/activities* von Shodor o. D.). Nach einem 3-2-1-Rückwärtszählen startet die Stoppuhr und alle beginnen, einen Ton zu singen, und versuchen, ihn möglichst lange zu halten.

Man kann die Anleitungen etwa so präzisieren: Alle singen einen Ton wie „ohhhhhh“ in der Mitte ihrer eigenen Stimmlage (das sollte die Ergebnisse untereinander besser vergleichbar machen und die Dauer fürs Durchhalten verlängern). Alle Studierenden haben den Auftrag, die Stoppuhr genau zu verfolgen und die Zeit zu notieren, ab wann sie nicht mehr in der Lage waren, den Ton fortzusetzen.

Kaum jemand wird über 18 Sekunden aushalten, sodass dieser Teil wenig Zeit in Anspruch nimmt. Nun wird man auf der Leinwand eine Tabellenkalkulation (oder eine andere Software) einblenden und die Da-

ten dort eintragen. Am einfachsten werden alle der Reihe nach ihre Zeit und die Position (sitzend oder stehend) laut ansagen. Wenn man mehr Diskretion will, so kann man die Daten (ohne Namen) auf einzelne Blätter eintragen und einsammeln lassen.

Um den statistischen Test auf Gleichheit der Mittelwerte durchzuführen, wird man sich erst einigen müssen, ob man einen *t*- oder einen *z*-Test (Gauß-Test) anwendet, und, ob man die Varianzen der beiden Populationen als gleich oder ungleich annimmt. Über die Berechnungen hinaus sollte man die Studierenden auch anregen, folgende Fragen zu erörtern:

- Ist das Experiment verblindet? Wenn ja, verblindet für die Probanden, für die Untersucher oder für beide?
- Was ändert sich, wenn man das Experiment so abändert, dass alle Versuchspersonen sowohl sitzend als auch stehend den Ton zu halten versuchen.
- Man soll die Auswirkungen einer dritten Versuchsbedingung untersuchen. Etwa: beim Singen die Arme ganz nach oben strecken; Singen nach 20 Hampelmann-Sprüngen. Dies führt dazu, dass wir mehr als zwei Mittelwerte auf Gleichheit testen, was zur Methode der Varianzanalyse überleitet.

Natürlich ist Singen eine ganz persönliche Form des Ausdrucks; das Schöne am Experiment ist aber, dass die Fähigkeiten zu singen, den Rhythmus zu halten etc., gar nicht gebraucht werden; es reicht, einen Ton einfach zu halten. Und alle singen parallel zueinander, sodass keiner wirklich hören kann, wenn es bei einzelnen gut oder weniger gut klingt (und, wenn die Ergebnisse per Zettel eingesammelt werden, hat kaum jemand einen Überblick, wer lange oder wer kurz ausgehalten hat).

Um das Experiment „sicher“ und „lustig“ zu gestalten, stelle ich mich selbst hin und zeige, wie man einen Ton hält und lasse die Stoppuhr mitlaufen. Und ich streiche hervor, dass jede Stimme im Chor untergehen wird und alle, die sich dennoch scheuen zu singen, das Experiment auch auslassen können. Auch könnte summen besser geeignet sein als singen („mmmmmm“ mit geschlossenen Lippen). Ich habe noch nicht erlebt, dass sich Einzelne verschlossen zeigen oder es gar ablehnen, teilzunehmen. Im Gegenteil, es scheint ein positiv besetztes gemeinsames Erleben zu sein, ungewöhnlich zwar für den Statistikerunterricht, aber angenehm. Und das Zusammenklängen der einzelnen Stimmen ergibt immer gute Schwingungen und interessante Harmonien.

5 Aktivität 4: Welche Hand ist schneller?

Hintergrund: Anderson-Cook & Sundar (2001a) beschreiben ein Klassenexperiment, in welchen die Reaktionszeiten der dominanten und der nicht-dominanten Hand miteinander verglichen werden. Das Experiment beginnt mit einer klaren 3-2-1-Startanweisung. Eine von neun Zellen leuchtet am Bildschirm auf und die Versuchsperson muss mit der gerade untersuchten Hand möglichst rasch die Computermaus auf diese Zelle steuern und die Zeit mit einem Klick auslösen. Die Zeit bis dahin wird auf Hundertstel genau gemessen.

Wenn es die Zeit erlaubt, kann man im Unterricht davor über die Prinzipien von Versuchsdesigns sprechen und den „ersten Ansatz“ im experimentellen Design im angesprochenen Artikel kritisch beleuchten lassen. Ist dagegen die Zeit knapp, so kann man gleich zur endgültigen Version des Applets (Anderson-Cook & Sundar 2001b) gehen.

Statistisches Thema: Test von Mittelwerten aus zwei abhängigen Stichproben.

Ablauf: Man beginnt das Experiment einfach mit der Projektion des Applets von Anderson-Cook & Sundar und führt persönlich mit jeder der beiden Hände vor, wie es funktioniert. Dabei erklärt man auch, was „dominant“ und „nicht-dominant“ bedeutet und wie die „Verkehrssampel“ sowie die Schaltflächen „Reset“ und „Start“ im Applet funktionieren.

Danach sollen einige Studierende den Test durchmachen (es reichen 5 oder 6 um die Begriffe rund um den besagten Test zu illustrieren, es können natürlich, sofern Zeit ist, mehr sein). Jeder kommt der Reihe nach zum Test und arbeitet am Applet. Jedes Datum braucht nur eine (!) Sekunde, sodass die Daten rasch erzeugt sind.

Mit der Schaltfläche „Compile Info“ werden die erzeugten Daten ausgewertet und der t -Wert des Tests auf Mittelwertunterschiede aus verbundenen Stichproben angegeben. Es bleibt, den p -Wert auszurechnen und damit das Ergebnis zu bewerten.

Literatur

Franklin, C. et al. (2007): Guidelines for assessment and instruction in statistics education (GAISE) Report. A pre-K-12 curriculum framework. Alexandria, VA: American Statistical Association. www.amstat.org/education/gaise/ (Zugriff: 8.8.2014).

Aliaga, M. et al. (2010): Guidelines for assessment and instruction in statistics education: College Report. Alexandria, VA: American Statistical Association. www.amstat.org/education/gaise/ (Zugriff: 8.8.2014).

Anderson-Cook, C. M.; Sundar, D.-R. (2001a): An active learning in-class demonstration of good experimental design. In: *Journal of Statistics Education* 9(1). www.amstat.org/publications/jse/v9n1/anderson-cook.html (Zugriff: 8.8.2014).

Anderson-Cook, C. M.; Sundar, D.-R. (2001b): Experimental Design Applets. In: *Journal of Statistics Education* 9(1). www.amstat.org/publications/jse/java/v9n1/anderson-cook/GoodExpDesignApplet.html (Zugriff: 8.8.2014).

Boundary Institute (o. D.): Got PSI? www.gotpsi.org (Zugriff: 8.8.2014).

Cheng, S. C.-T. (1991): The Tao of voice: a new East-West approach to transforming the singing and speaking voice. Rochester, VT: Destiny Books.

Kaya, N.; Burgess, B. (2007): Territoriality: seat preferences in different types of classroom arrangements. In: *Environment and Behavior* 39(6), S. 859–876.

Lesser, L. M.; Glickman, M. E. (2009): Using magic in the teaching of probability and statistics. In: *Model Assisted Statistics and Applications* 4(4), S. 265–274. math.bu.edu/people/mg/research/Magic-Statistics-post.pdf (Zugriff: 8.8.2014).

Lesser, L. M.; Pearl, D. K. (2008): Functional fun in statistics teaching: Resources, research, and recommendations. In: *Journal of Statistics Education* 16(3), 10 S. www.amstat.org/publications/jse/v16n3/lesser.pdf (Zugriff: 8.8.2014).

Shodor (o. D.): Shodor – a national resource for computational science resources. www.shodor.org/activities/ (Zugriff: 8.8.2014).

Utts, J. M. (2005): Seeing through statistics. 3rd edition. Belmont, CA: Thomson Brooks/Cole.

Anschrift des Verfassers

Lawrence M. Lesser
The University of Texas
Mathematical Sciences Department
500 West University Avenue
El Paso, TX 79902, USA
lesser@utep.edu