

Randomisierungstests als Einstieg in die Inferenzstatistik für Grundschullehramtsstudierende

Schlussfolgerungen aus Daten zu ziehen, ist eine Fähigkeit, die im Alltag häufig benötigt wird. Beinahe täglich begegnen uns Situationen, in denen Entscheidungen getroffen werden müssen, die auf Daten beruhen. Genauso ist ein reflektierter Umgang mit Medienberichten nötig, denn häufig hat „eine Studie gezeigt, dass ...“ oder wurde geprüft, dass „ein Effekt von A an B liegt“. In den Medien wird jedoch häufig nicht erwähnt, dass die so dargestellten Ergebnisse und Interpretationen keineswegs sicher sind, sondern auf Praktiken der Inferenzstatistik beruhen und Unsicherheiten beinhalten. Somit sind Kenntnisse in der Inferenzstatistik nötig, um als informierter Bürger datenbasierte Entscheidungen treffen, bzw. beurteilen zu können (Biehler & Engel, 2015). Allerdings sind bezüglich des Lernens und Verwendens inferenzstatistischer Methoden viele Schwierigkeiten bekannt, und allgemein gilt die Logik der Inferenzstatistik als herausfordernd (z. B. Garfield & Ben-Zvi, 2008).

Statistisches Schlussfolgern mit Randomisierungstests

Ein Zugang zur Inferenzstatistik, der als besonders gut geeignet angesehen wird, um in die Logik des inferentiellen Denkens einzuführen, bieten Randomisierungstests (Cobb, 2007). Randomisierungstests gehören zu den parameterfreien Verfahren, die auf computerbasierten Simulationen beruhen und auch deshalb als leicht zugänglich angesehen werden (Budgett, Pfannkuch, Regan, & Wild, 2012). Ein Randomisierungstest ist (heutzutage) ein computerbasiertes Verfahren zur statistischen Analyse von Daten, deren Verteilungseigenschaften nicht bekannt sind. Es ist ein Verfahren, das besonders für kleine Stichproben oder auch abhängige Daten geeignet ist und deshalb gerade in medizinischen, pädagogischen oder psychologischen Bereichen Anwendung findet, in denen häufig mit kleinen Stichproben gearbeitet wird. Beim Durchführen eines tatsächlichen Experiments werden die Versuchseinheiten zufällig auf zwei (oder mehr) Gruppen aufgeteilt und diese werden dann unterschiedlichen Behandlungen (treatments) ausgesetzt. Bei der Analyse der Ergebnisse der beiden Gruppen werden bestimmte Kennzahlen gemessen und verglichen, beispielsweise als Differenz der Mittelwerte beider Gruppen.

Die Grundidee des Randomisierungstests ist es, dass alle denkbaren Möglichkeiten von zufälligen Gruppenzuordnungen unter Beibehaltung der erreichten Werte durchgespielt werden und die kumulative Wahrscheinlichkeit

bestimmt wird, mit der ein Ergebnis eintritt, das genauso extrem wie oder extremer als ein zuvor beobachtetes Ergebnis ist. Schlussfolgerungen bei einem signifikanten Ergebnis sind dann möglich bezüglich der Ursache, also z.B. bezogen auf die durchgeführte Behandlung, nicht aber i.A. bezüglich der Population (Ramsey & Shafer, 2013).

Für den anglo-amerikanischen Raum existieren sowohl Curricula als auch kurze Unterrichtsvorschläge (z. B. Budgett et al., 2012), um mittels Randomisierungstests in die Logik der Inferenzstatistik einzuführen. Im Rahmen meiner Dissertation habe ich eine kurze Lerneinheit „Inferenzstatistik mit Randomisierungstests“ nach dem Design-Based Research Ansatz für Grundschullehrantsstudierende entwickelt (Podworny, 2018) und Argumentationsprozesse von Lernenden untersucht, die einen Randomisierungstest durchführen.

Studie zur Argumentation von Lernenden bei der Durchführung eines Randomisierungstests

Um zu untersuchen, welche Faktoren für gelingendes kollaboratives Arbeiten der Lernenden bei der Durchführung eines Randomisierungstests unter Verwendung der Software TinkerPlots festgestellt werden können und welche Problemstellen in diesem Prozess auftreten, habe ich eine qualitative Studie mit Studierenden des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung der Universität Paderborn durchgeführt.

Aufgabe: Fischöl und Blutdruck – eine Studie

Menschen mit Bluthochdruck haben eine höhere Wahrscheinlichkeit, einen Herzinfarkt zu erleiden. Deshalb suchen Forscher nach einem Weg, den Blutdruck von Bluthochdruckpatienten zu senken. Dazu wurde diese Studie durchgeführt. Vor der Durchführung dieser Studie vermuteten Forscher, dass Personen, die an einer Fischöl-Diät teilnehmen, dazu neigen, dass ihr Blutdruck stärker gesenkt wird, als bei denjenigen, die eine normale Diät halten. Für die Studie wählten die Forscher 14 männliche Freiwillige mit hohem Blockdruck aus und teilten sie zufällig in zwei Gruppen ein. Die erste Gruppe machte vier Wochen lang eine „Fischöl-Diät“ und die zweite Gruppe machte eine „normale Öl-Diät“ für vier Wochen. Die Behandlungsgruppe ist also die Fischöl-Diät Gruppe und die Kontrollgruppe ist die normale Öl-Diät Gruppe. Von jedem Teilnehmer wurde am Anfang und am Ende der Studie der Blutdruck gemessen. Die daraus resultierenden Blutdruckreduktionen(-senkungen) in Millimeter Quecksilber waren:

Fischöl-Diät	8	12	10	14	2	0	0
Normales Öl-Diät	-6	0	1	2	-3	-4	2

Abbildung 1: Punktdiagramm und Boxplot mit Differenz der Mittelwerte zur Fischöl Studie

Die beobachteten Daten sind in Abbildung 1 dargestellt und zeigen, dass die Senkungen der Blutdruckwerte in der „Fischöl“ Gruppe tendenziell größer sind als die der „normale Öl“ Gruppe.

Was lässt sich hier nun schließen?

Abbildung 1: Aufgabe der Studie

An der Studie haben sechs Studierende jeweils zu zweit teilgenommen, und einen Randomisierungstest zur Aufgabe in Abb.1 durchgeführt. Für die computerbasierte Simulation stand die Software TinkerPlots zur Verfügung.

Alle Teilnehmer der Studie hatten zuvor an der Lerneinheit „Inferenzstatistik mit Randomisierungstests“ teilgenommen, die in einer fachlichen Lehrveranstaltung in zwei Vorlesungen und einer Übung am Ende des Sommersemesters 2017 implementiert wurde.

Die durchgeführten Handlungen (mit und ohne Software) und die sprachlichen Äußerungen während der Bearbeitung der Aufgabe wurden aufgenommen und transkribiert.

Analyse und Auswertung

Mit dem Ansatz der Gesprächsanalyse (Deppermann, 2008) wurden Gesprächsstrukturen herausgearbeitet und 15 Interaktionseinheiten identifiziert. Die Bearbeitungsprozesse der Studierenden wurden bezüglich des fachlichen Inhalts ausgeweitet auf Softwareaktivitäten mit Mitteln der Interaktionsanalyse (Krummheuer & Naujok, 1999) analysiert. Dazu wurden die Bearbeitungen hinsichtlich der neun für einen Randomisierungstest nötigen Konzepte *Forschungsfrage aufstellen, zufällige Zuordnung erklären, beobachtete Daten analysieren, mögliche Erklärungen nennen, Nullmodell aufstellen, Simulationsmodell aufstellen, Teststatistik und Referenzverteilung darstellen, P-Wert ermitteln* und *Schlussfolgerungen ziehen* analysiert.

Ergebnisse

Unter Laborbedingungen wurde die Studie zur kollaborativen Bearbeitung eines Randomisierungstests durchgeführt. Ein positives Ergebnis ist, dass alle drei Paare die neun Konzepte adressiert haben. Die Interaktionseinheiten waren dabei mit wenigen Ausnahmen geprägt von kurzen Turns und grammatikalisch unvollständigen Sätzen, durch die jedoch kollaborativ gearbeitet wurde und somit Wissen gemeinsam formuliert wurde. Das kollaborative Arbeiten hat sich hier als nützlich herausgestellt, um die Aufgabe gemeinsam zu bearbeiten. Bezüglich des Arbeitens mit der Software konnte festgestellt werden, dass alle Paare Hilfe bezüglich der Bedienung benötigten, trotzdem jedoch häufig vorausschauendes Planen zu beobachten war. Schwierigkeiten konnten auf technischer Ebene beobachtet werden, das Verständnis der einzelnen Schritte in Bezug auf einen Randomisierungstest war jedoch vorhanden.

Auf inhaltlicher Ebene wurden mehrere Problemstellen identifiziert. So hatte ein Paar beispielsweise Schwierigkeiten, mögliche Erklärungen für die beobachteten Unterschiede zu nennen. Bei zwei Paaren zeigte sich, dass sie

versuchten, mithilfe des gewonnenen P-Werts Angaben zur Hypothesenwahrscheinlichkeit zu machen, wodurch sich dieses Wissen als fragil herausstellte. Bezüglich der weiteren Konzepte zeigte sich jedoch, dass zu diesen weitestgehend zufriedenstellend argumentiert werden konnte.

Ausblick

Es hat sich auf der einen Seite gezeigt, dass die sechs Studierenden mit dem in der kurzen, sechs Semesterwochenstunden umfassenden Lerneinheit erworbenen Wissen mit Hilfestellung in Bezug auf die Softwarebenutzung in der Lage waren, einen Randomisierungstest durchzuführen. Hierzu haben alle Teilnehmer im Interview geäußert, dass die in der Lerneinheit durchgeführte händische Durchführung eines Randomisierungstests hilfreich war, um den Prozess des Randomisierens zu verstehen. Dies steht in Einklang mit Ergebnissen von Budgett et al. (2012). Auf der anderen Seite hat sich gezeigt, dass das Beherrschen der Software durch den Lernenden notwendig ist, um sich auf die Logik des inferentiellen Schließens konzentrieren zu können. Aufgrund der gewonnenen Erkenntnisse lässt sich die zugrunde liegende Lerneinheit als erfolgreich einstufen.

Literatur

- Biehler, R., & Engel, J. (2015). Stochastik: Leitidee Daten und Zufall. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thime, & H. G. Weigand (Eds.), *Handbuch der Mathematikdidaktik* (pp. 221-251). Berlin, Heidelberg: Springer.
- Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. (2012). *Dynamic visualizations for inference*. Paper presented at the The International Association for Statistical Education Roundtable Conference: Technology in statistics education: Virtualities and Realities, Cebu City, The Philippines.
- Cobb, G. (2007). The Introductory Statistics Course: A Ptolemaic Curriculum? *Technology Innovations in Statistics Education*, 1(1), 1-15. doi:<https://escholarship.org/uc/item/6hb3k0nz>
- Deppermann, A. (2008). *Gespräche analysieren. Eine Einführung*. Wiesbaden: Verlag für Sozialwissenschaften.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer Science+Business Media.
- Krummheuer, G., & Naujok, N. (1999). *Grundlagen und Beispiele interpretativer Unterrichtsforschung*. Opladen: Leske+Budrich.
- Podworny, S. (2018). *Simulationen und Randomisierungstests mit der Software Tinker-Plots. Theoretische Werkzeuganalyse zur stochastischen Simulation und explorative Fallstudie zum statistischen Schließen mit Randomisierungstests*. Paderborn: Universität Paderborn.
- Ramsey, F. L., & Shafer, D. W. (2013). *The Statistical Sleuth. A Course in Methods of Data Analysis*. Boston, Massachusetts: Cengage Learning.