Randomisierungstests mit TinkerPlots

Rolf Biehler & Daniel Frischemeier, Paderborn

Zusammenfassung: *Wenn man Verteilungen eines numerischen Merkmals in gruppierten Datensätzen vergleicht, die innerhalb von Experimenten erhoben worden sind und dort Unterschiede (z. B. bezüglich der Mittelwerte) herausarbeitet, stellt man sich oft die Frage, ob in diesem Fall der Unterschied zwischen den Mittelwerten zufällig entstanden ist oder das eine andere Ursache hat. Eine Möglichkeit, sich dieser Frage zu nähern, ist das Durchführen eines Randomisierungstests. Da Randomisierungstests im Allgemeinen anspruchsvoll und auch rechenaufwändig sind, möchten wir im Folgenden einen kleinen Einblick geben, wie bereits Schülerinnen und Schüler ab der Sekundarstufe I das Durchführen von Randomisierungstests unterstützt durch die Software TinkerPlots erfahren können.*

1 Einleitende Bemerkungen

Das Erheben und Analysieren von Daten und damit einhergehend das Vergleichen von Verteilungen eines numerischen Merkmals stellt eine wichtige Aktivität in der Statistik dar und wird auch curricular in den Bildungsstandards (KMK 2004), den Empfehlungen des AK Stochastik (2003) sowie in den NCTM-Standards (2000) gefordert. Kommt es zu den besagten Verteilungsvergleichen wird in der didaktischen Literatur verschiedentlich vorgeschlagen, direkt hieran Randomisierungstests anzuschließen, um eine Brücke zur Inferenzstatistik zu bauen (vgl. z. B. Watson & Moritz (1999) oder Makar & Confrey (2002)). Wie man Randomisierungstests in diesen Kontexten einsetzen kann, z. B. am Ende der Sekundarstufe I, um Ergebnisse aus Experimenten (Sprungweiten von leichten Papierfröschen vs. Sprungweiten von schweren Papierfröschen) zu verallgemeinern, führt u. a. Vogel (2009) aus. Die Durchführung von Randomisierungstests im Schulunterricht als Weiterführung von Verteilungsvergleichen wird z. B. in Biehler, Kombrink und Schweynoch (2003, 20) aufgegriffen. Plädoyers für die Thematisierung von Randomisierungstests schon zu Beginn von Statistikkursen existieren u. a. von Cobb (2007) und Rossman (2008). Es bleibt zu bemerken, dass die letztgenannten Plädoyers sich vornehmlich auf die Thematisierung dieser Inhalte im tertiären Sektor beziehen. Cobb (2007) betont, dass einerseits die Umsetzung eines solchen Tests ohne eine Unterstützung durch adäquate Software schwierig sei, räumt andererseits aber auch ein, dass sich die Durchführung eines solchen Tests auch nicht zu sehr auf die Nutzung einer Software konzentrieren sollte. Mit den Randomisierungstests wird es möglich, Unterschiede zwischen zwei Stichproben auf Signifikanz zu testen. Die vergleichbaren „klassischen Verfahren", wie z. B. der 2-Stichproben-t-Test sind nicht Teil des üblichen Schulcurriculums und setzen einen erheblichen theoretischen und algebraischen Aufwand voraus.

Eine Anwendung von Randomisierungstests findet sich z. B. in klinischen Studien (siehe z. B. Erickson, 2006). k Patienten werden zufällig (ohne Zurücklegen) in eine Kontroll- und eine Experimentalgruppe eingeteilt. Die Experimentalgruppe bekommt ein „Treatment", dessen mögliche Wirkung mit einem Kriterium X gemessen wird. Man berechnet z. B. die beobachtete Differenz $X_{exp} - X_{kontroll}$. Wenn das Treatment keinen Effekt herbeiführt, ist die hier ermittelte Differenz $X_{exp} - X_{kontroll}$ ein Produkt der zufälligen Einteilung in Kontroll- und Experimentalgruppe. Wie wahrscheinlich ist es, dass man eine solche Differenz oder eine noch extremere durch die zufällige Gruppeneinteilung bekommt? Diese Wahrscheinlichkeit ermittelt man, indem man alle möglichen zufälligen Gruppeneinteilungen durchspielt und jeweils die Differenz notiert. Man greift dazu von den k Objekten jeweils eine Teilmenge mit k/2 Elementen zufällig heraus. Die Objekte in dieser Teilmenge werden mit dem Etikett „Pseudo-Experimentalgruppe" versehen. Die übrigen Objekte werden mit dem Etikett „Pseudo-Kontrollgruppe" versehen. Dabei wird man gewisse Unterschiede zwischen den Mittelwerten in den beiden Gruppen erwarten. Als Vergleich führt man die Gruppeneinteilung sehr oft zufällig durch und bekommt damit eine Verteilung der gesammelten Werte des Merkmals $X_{pseudoexp} - X_{pseudokontroll}$, mit der man das beobachtete Ergebnis im realen Experiment vergleichen kann. Als Kriterium wird der P-Wert der real beobachteten Differenz auf der Basis der so genannten „Randomisierungsverteilung" gewählt. Stellt man eine signifikante Abweichung fest, so kann man das kausal auf das Treatment zurückführen. Die Methode ist auch anwendbar, wenn die Versuchspersonen nicht zufällig aus einer größeren Grundgesamtheit gewählt wurden, dann aber ist eine Verallgemeinerung über die am Experiment teilgenommenen Personen/Objekte mit statistischen Mitteln nicht möglich, sondern kann ggf. nur aus den entsprechenden sachbezogenen Theorien begründet werden.

2 Ein Anwendungsbeispiel

Ein paradigmatisches Beispiel einer solchen klinischen Studie findet sich z. B. in Lock, Lock, Lock, Lock & Lock (2013, 240 – Übersetzung: Daniel Frischemeier):

„Viele Menschen sind überzeugt davon, dass sie morgens einen Kaffee oder eine Form von Koffeinkonsum benötigen, um fit für den bevorstehenden Tag zu sein. Die Effekte von Koffein auf den menschlichen Körper wurden und werden intensiv studiert. In einem der vielen Experimente[1] trainierten Forscher mit einer Gruppe von männlichen Collegestudenten so oft wie möglich mit dem Zeigefinger auf den Tisch zu tippen. Die besagten Collegestudenten wurden nach dem Training, aber vor dem eigentlichen Experiment, zufällig in zwei Gruppen zu jeweils zehn Personen eingeteilt. Die Studenten in Gruppe 1 bekamen Kaffee mit 200 mg Koffein, die Studenten in Gruppe 2 bekamen entkoffeinierten Kaffee (0 mg Koffein). Nach zwei Stunden wurde jeder Student aufgefordert, so schnell wie möglich mit dem Zeigefinger auf den Tisch zu tippen. Die Anzahl der Fingertipps pro Minute wurde gezählt und notiert (siehe Abb. 1).

Koffein	246	248	250	252	248	250	246	248	245	250
Kein Koffein	242	245	244	248	247	248	242	244	246	242

Abb. 1: Daten aus dem Experiment

Die Probanden wussten nicht, ob ihre Getränke Koffein enthielten oder nicht. Auch die Beobachter, die die Messung vornahmen und die Messungen dokumentierten, wussten dies nicht. Nur der auswertende Statistiker hatte die entsprechenden Informationen und konnte die Gruppenzugehörigkeit und die gemessenen Tippraten zuordnen. Das Ziel des Experiments war zu entscheiden, ob Koffein eine Erhöhung der durchschnittlichen Tipprate bewirkt."

Der Mittelwert der Koffeingruppe ist 248,3 Tipps/Minute, der der Nicht-Koffein-Gruppe ist 244,8 Tipps/Minute – es besteht also eine Differenz von 3,5 Tipps/Minute. Wie ist diese Differenz zu beurteilen? Wenn das Koffein keinen Effekt hat, dann spielt die Gruppeneinteilung keine Rolle für die erhaltenen Messungen.

Wie wir der Tabelle (Abb. 1) entnehmen können, wurden von 20 Personen Daten erhoben. Von diesen 20 Personen wurden 10 Personen zufällig herausgegriffen und mit dem Etikett „Koffeingruppe" versehen. Die anderen 10 Personen wurden mit dem Etikett „Nicht-Koffein-Gruppe" versehen. Dabei wird man ja immer gewisse Unterschiede zwischen den Mittelwerten in den beiden 10er Gruppen erwarten. Aber welche Unterschiede kommen bei rein zufälliger Gruppeneinteilung vor? Hierüber kann man nur Wahrscheinlichkeitsaussagen machen.

Wir interessieren uns dafür, wie wahrscheinlich es ist, bei rein zufälliger Gruppeneinteilung eine Differenz von 3,5 oder größer zu bekommen. Ist diese Wahrscheinlichkeit gering, z. B. kleiner gleich 5 %, so zweifeln wir die Annahme an, dass der Unterschied zwischen Koffeingruppe und Nichtkoffeingruppe durch eine zufällige Gruppeneinteilung ohne Wirkung auf die Fingertipps zustande gekommen ist.

Ein erster handlungsbasierter Zugang hierzu besteht darin, die 20 Messwerte auf 20 gleiche Kärtchen zu schreiben und verdeckt 10 zu ziehen und aufzudecken. Dann wird der Mittelwert der aufgedeckten 10 Kärtchen berechnet, sowie der Mittelwert der 10 zunächst nicht aufgedeckten Kärtchen. Dann wiederholt man den Prozess und plottet die so erhaltenen Differenzen in ein Diagramm, dort entsteht dann nach und nach eine Verteilung. Man kann dann sehen, ob die Differenz von 3,5 in einem extremen Randbereich der Verteilung liegt oder eher in der Mitte der Verteilung. Um durch die relative Häufigkeit, mit der die Differenz von 3,5 überschritten wird, die Wahrscheinlichkeit angenähert schätzen zu können, ist es sinnvoll im Sinne des $1/\sqrt{n}$-Gesetzes (vgl. Biehler & Prömmel, 2013), die Rerandomisierungsanzahl zu erhöhen, bzw. hoch genug anzusetzen. Diesen realen „Rerandomisierungsprozess“ kann man relativ einfach in die Software TinkerPlots übertragen (Konold & Miller, 2011), die wir im Folgenden kurz vorstellen werden.

3 Die Software TinkerPlots

Die Software TinkerPlots ist eine von Cliff Konold und Craig Miller entwickelte Datenanalysesoftware, die ursprünglich für den Einsatz in den Klassen 3 bis 8 vorgesehen war, aber mittlerweile bis zum Collegeniveau und in der Lehrerausbildung eingesetzt wird (Biehler, Ben-Zvi, Bakker & Makar, 2013). Inzwischen existiert auch eine deutsche Version der Software (Adaption: Rolf Biehler, Daniel Frischemeier und Susanne Podworny) und kann unter https://www.tinkerplots.com/get (unter Auswahl der Sprache „Deutsch“ im Menü „Language“) heruntergeladen werden.

In einer aktualisierten Version 2.0 wurde ein Tool („Zufallsmaschine“) implementiert, welches es ermöglicht mithilfe diverser Bauteile (Box, Zähler, Kreisel, etc.) Simulationen von Zufallsexperimenten zu modellieren und zu simulieren. Details zum Datenanalysepotential von TinkerPlots finden sich in Biehler (2007) sowie in Biehler & Frischemeier (2015). Details zum Simulationspotential können in Biehler, Frischemeier und Podworny (2015a) sowie in Biehler, Frischemeier & Podworny (2016) nachgelesen werden. Beispiele für den Einsatz der Software TinkerPlots beim Durchführen von Randomisierungstests finden sich in z. B. in Watson (2013).

Mit ihrer visualisierten Zufallsmaschine scheint TinkerPlots unseres Erachtens in zweierlei Hinsicht für die Durchführung eines solchen Tests in besonderem Maße geeignet zu sein: Zum einen weil die Zufallsmaschine die Modellierung und Auswertung eines stochastischen Zufallsexperiments ohne (wie es z. B. bei Tabellenkalkulationsprogrammen der Fall wäre) Kenntnisse bestimmter Formeln und Programmierbefehlen ermöglicht, zum anderen aufgrund der Visualisierung des Modells und des Ablaufs einer Simulation, die es dem Lernenden ermöglicht, den Vorgang sehr gut nachzuvollziehen.

4 Randomisierungstests mit der Software TinkerPlots

Wir wollen an dem oben aufgeworfenen Beispiel aus Lock et al. (2013, 240) das Potential der Software aufzeigen, einen solchen Randomisierungstest durchzuführen und für Schülerinnen und Schüler zugänglich zu machen. Nach der Eingabe der Daten in die Software TinkerPlots kann man mithilfe eines Graphen und dem Einblenden der arithmetischen Mittelwerte (siehe Abb. 2) die Mittelwert-Unterschiede zwischen beiden Verteilungen berechnen.

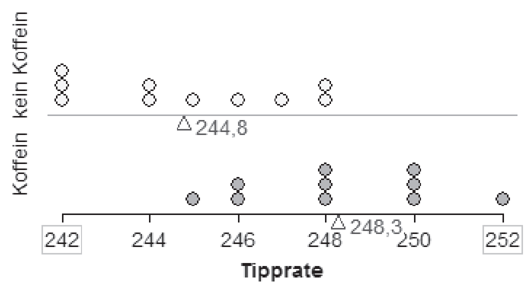

Abb. 2: Vergleich der Verteilungen des Merkmals „Tipprate“ nach Koffeinkonsum

Wir sehen, dass bei den Koffein-Konsumenten die Tipprate (248,3 Tipps/Minute) im Durchschnitt um 3,5 Tipps/Minute größer ist als bei denen, die den entkoffeinierten Kaffee getrunken haben (244,8 Tipps/Minute).

Wir „simulieren“ nun mit TinkerPlots, dass der Unterschied zwischen Koffein- und Nicht-Koffein-Konsumenten in dem hier vorliegenden Experiment zufällig entstanden ist.

Dazu nehmen wir das Bauteil Zähler (Abb. 3, links) und beschriften diesen mit den 20 Ausprägungen des Merkmals „Tipprate“ und des Merkmals „Gruppe“ aus dem Datensatz (Abb. 1).[2] Dieses lässt sich in TinkerPlots durch Copy&Paste aus der ursprünglichen Datentabelle leicht realisieren. Der Angabe über dem Zähler in Abb. 3 („20 Fälle, 2 Merkmale“) können wir entnehmen, dass im Zähler nun unsere 20 Fälle aus dem Experiment (Abb. 1) repräsentiert sind. Dabei werden im Zähler nur die Ausprägungen des Merkmals „Tipprate“ angezeigt, die Ausprägungen des Merkmals „Gruppe“ sind zwar im Zähler enthalten, werden aber nicht angezeigt.

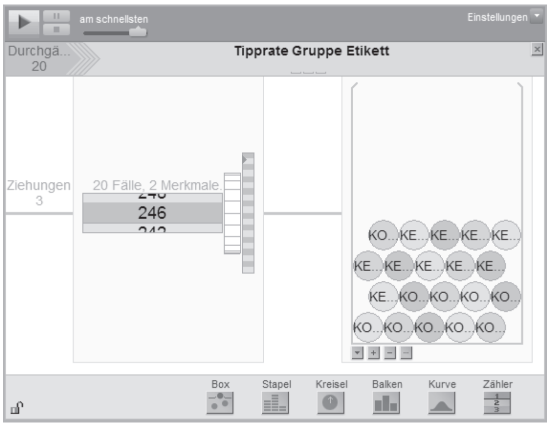

Abb. 3: Zufallsmaschine in TinkerPlots zur Realisierung des Randomisierungstests (auf dem Zähler befinden sich die Werte des Merkmals „Tipprate“ und des Merkmals „Gruppe“ (hier nicht sichtbar), die die echten Daten aus dem Experiment repräsentieren; in der Urne rechts befinden sich die 20 Kugeln mit jeweils 10 Etiketten „KOFFEIN“ und 10 Etiketten „KEIN KOFFEIN“)

Wir legen dann 20 Kugeln (10mal „KEIN KOFFEIN“ und 10mal „KOFFEIN“) in eine Urne (Abb. 3, rechts). Wir nennen diese Urne „Etikett“ und beschriften die Ausprägungen dieses Merkmals in Großbuchstaben, um sie von den Ausprägungen im vorliegenden Experiment unterscheiden zu können.

Nun starten wir die Zufallsmaschine: Das Bauteil „Zähler“ in TinkerPlots gibt sukzessive nacheinander die Werte des Merkmals „Tipprate“ und des Merkmals „Gruppe“ aus der Tabelle in Abb. 1 wieder. Jedem dieser Wertepaare wird dann nacheinander zufällig aus der rechten Urne ein gezogenes Etikett verpasst (ohne Zurücklegen). Dieser Vorgang wird 20mal wiederholt, bis keine Kugel mehr in der rechten Urne liegt. Wichtig ist, dass allen Messwerten mit ihren Eingruppierungen zufällig ein Etikett zugeordnet wird, das entspricht der zufälligen Gruppeneinteilung, wenn kein Effekt des Treatments vorliegt. Die Zufallsmaschine in TinkerPlots visualisiert diesen Vorgang für den Benutzer anschaulich: In jedem Durchgang wird einer „Person“ mit einer Tipprate zufällig ein Etikett zugelost. Durch Veränderung der Geschwindigkeit an der Zufallsmaschine kann dieser Prozess beschleunigt oder verlangsamt werden.

TinkerPlots versieht also die Ausprägungen des Merkmals „Tipprate“ und des Merkmals „Gruppe“ mit unserem „Etikett“ und dokumentiert die „neuen“ Zuordnungen in einer Tabelle (Abb. 4, siehe rechte Spalte). Mit Blick auf die Tabelle sehen wir in der linken Spalte die Ausprägungen des Merkmals „Tipprate“ (Anzahl Tipps/Minute) und in der mittleren Spalte die Ausprägungen des Merkmals „Gruppe“, wie sie im Experiment vorhanden sind. In der Spalte rechts sehen wir die zufällige Zuordnung der Etikettierungen zu den einzelnen Probanden aus dem Experiment.

Konkret stellen wir, wenn wir nun die einzelnen Zeilen der Tabelle in Abbildung 4 betrachten, fest, dass zum Beispiel die Person 1 (die in Wirklichkeit in der „Koffein“-Gruppe ist) das Etikett „KEIN KOFFEIN“, die Person 2 (die in Wirklichkeit in der „Koffein“-Gruppe ist) das Etikett „KEIN KOFFEIN“ und die Person 3 (die in Wirklichkeit in der „Koffein“-Gruppe ist ist) das Etikett „KOFFEIN“ zugeordnet bekommen hat, usw.

Ergebnisse von Zufallsmaschine 1 — Einstellungen

	Tipprate	Gruppe	Etikett
1	246	Koffein	KEIN KOFFEIN
2	248	Koffein	KEIN KOFFEIN
3	250	Koffein	KOFFEIN
4	252	Koffein	KEIN KOFFEIN
5	248	Koffein	KEIN KOFFEIN
6	250	Koffein	KOFFEIN
7	246	Koffein	KEIN KOFFEIN
8	248	Koffein	KOFFEIN
9	245	Koffein	KOFFEIN
10	250	Koffein	KOFFEIN
11	242	kein Koffein	KOFFEIN
12	245	kein Koffein	KEIN KOFFEIN
13	244	kein Koffein	KEIN KOFFEIN
14	248	kein Koffein	KOFFEIN
15	247	kein Koffein	KOFFEIN
16	248	kein Koffein	KEIN KOFFEIN
17	242	kein Koffein	KOFFEIN
18	244	kein Koffein	KEIN KOFFEIN
19	246	kein Koffein	KOFFEIN
20	242	kein Koffein	KEIN KOFFEIN

Abb. 4: Tabelle mit den Ergebnissen der Randomisierung

Das Ergebnis unserer Randomisierung können wir in Abbildung 5 sehen, indem wir nun die Verteilungen des Merkmals „Tipprate“ unterschieden nach der Etikettierung betrachten.

Mithilfe des Lineals (Pfeil in Abbildung 5 und 6) in TinkerPlots lassen sich nun die Differenzen der arithmetischen Mittelwerte der beiden Verteilungen berechnen. Hier muss man lediglich das Lineal an beiden Mittelwerten anlegen: TinkerPlots berechnet dann automatisch die jeweilige Differenz.

In der Situation, die zu Abbildung 4 gehört, beträgt der Unterschied der arithmetischen Mittelwerte beider Verteilungen 0,5 (Abb. 5). Das heißt, dass in diesem Fall die Personen mit dem Etikett „KOFFEIN“ durchschnittlich eine um 0,5 höhere Tipprate haben als die Personen mit dem Etikett „KEIN KOFFEIN“.

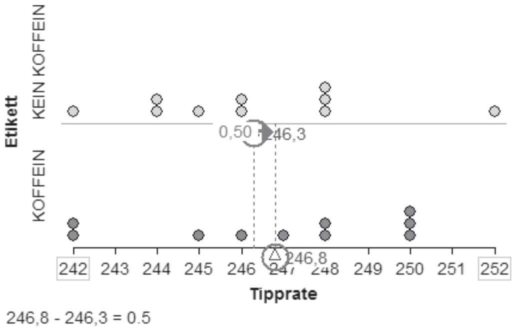

Abb. 5: Vergleich zweier „zufälliger“ Verteilungen nach einem Etikettierungs-Prozess (Randomisierungsprozess)

Diesen Prozess können wir in TinkerPlots per Knopfdruck wiederholen.

Nach einer Wiederholung der Randomisierung beträgt der Unterschied der beiden arithmetischen Mittelwerte beider Verteilungen 1,1 (Abb. 6). Das heißt hier haben die Personen mit dem Etikett „KOFFEIN“ durchschnittlich eine um ca. 1,1 (Anzahl/Minute) höhere Tipprate als die mit dem Etikett „KEIN KOFFEIN“.

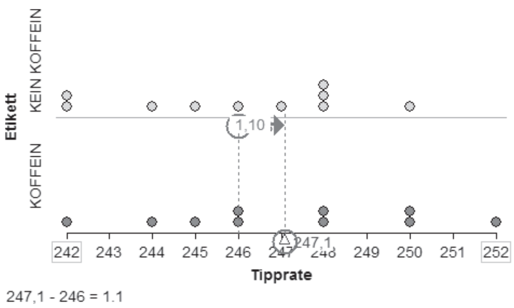

Abb. 6: Vergleich zweier „zufälliger“ Verteilungen nach einem weiteren Etikettierungs-Prozess (Randomisierungsprozess)

Diesen Prozess gilt es nun sehr oft zu wiederholen und die Unterschiede der arithmetischen Mittelwerte beider Verteilungen zu dokumentieren. TinkerPlots bietet hier eine Funktion „Messgrößen sammeln“ an, die es ermöglichst sogenannte Messgrößen verschiedener Kennzahlen einer Graphik zu sammeln. In diesem Fall soll der Unterschied der arithmetischen Mittelwerte in den beiden durch das Merkmal „Etikett“ getrennten Teilgruppen gesammelt werden. Das geht technisch sehr einfach, man wählt zunächst den Wert der Differenz in der Graphik an. Wenn man nun den Knopf „Messgrößen sammeln“ betätigt, wird eine Wiederholung der Simulation erzeugt und der neue Wert der Differenz in einer Tabelle (siehe Abb. 7) protokolliert. Mit einem weiteren Klick kann man eine N-fache Wiederholung dieser Simulation mit automatischem Protokollieren der variierenden Ergebnisse initiieren. Wir führen mithilfe von TinkerPlots 5000 dieser Randomisierungsvorgänge durch. Die automatisch dokumentierten 5000 Werte der Messgröße „Differenz“ werden in TinkerPlots in einer weiteren Tabelle (Abb. 7) dokumentiert. Wir können die Tabelle in Abbildung 7 nun so lesen: In der ersten Zeile finden wir das Ergebnis der ersten Randomisierung („die Personen mit dem Etikett KOFFEIN haben durchschnittlich eine um 0,5 höhere Tipprate als die Personen mit dem Etikett KEIN KOFFEIN“). In der zweiten Zeile finden wir das Ergebnis der zweiten Randomisierung („die Personen mit dem Etikett „KOFFEIN“ haben durchschnittlich eine um 1,1 höhere Tipprate als die Personen mit dem Etikett KEIN KOFFEIN“). In der dritten Zeile ist es nun umgekehrt, nun haben die Personen mit dem Etikett „KOFFEIN“ durchschnittlich eine um 3,3 geringere Tipprate als diejenigen mit dem Etikett „KEIN KOFFEIN“, usw.

Messgrößen der Erg… Sammeln 5000 Einstellungen

	Abstand_Tipprate	<neu>
1	0,5	
2	1,1	
3	-3,3	
4	-2,1	
5	0,9	
6	-0,3	
7	1,1	
8	-2,1	
9	-0,1	
10	0,5	
11	0,5	
12	-1,1	
13	1,1	
14	-0,1	
15	0,7	
16	1,7	
17	-0,7	

Abb. 7: gesammelte Messgrößen (Unterschiede der arithmetischen Mittelwerte der Verteilungen in den beiden durch „Etikett“ definierten Teilgruppen) in TinkerPlots

Diese Verteilung der Messgrößen lässt sich in TinkerPlots (siehe Abb. 8) visualisieren. Wir bezeichnen die Graphik kurz als Referenzgraphik. Wir wollen nun die Wahrscheinlichkeit bestimmen, dass durch zufällige Etikettierung ein Unterschied von 3,5 Tipps/Minute oder mehr entsteht.

Das ist der so genannte P-Wert. Dazu bestimmen wir den Anteil (die relative Häufigkeit) der Messgrößen, die größer oder gleich dem Unterschied in dem uns vorliegendem Experiment (= 3,5) sind. Dieses lässt sich in TinkerPlots durch sogenannte Einteiler (grauer Bereich in Abb. 8) leicht realisieren. Man kann über diese ein bestimmtes Intervall definieren, in dem dann die relativen Häufigkeiten der sich in dem

Intervall befindenden Fälle bestimmt werden. Die Wahrscheinlichkeit dafür, dass der Unterschied des arithmetischen Mittels der Verteilung des Merkmals „Tipprate" größer oder gleich dem im Experiment beobachteten Unterschied von 3,5 Tipps/Minute ist, unter der Voraussetzung, dass die Annahme („die Gruppenzugehörigkeit Koffein vs. kein Koffein hat keinen Einfluss auf die Fingertipprate") gilt, kann somit auf ca. 0,0038 geschätzt werden. Wenn man die aus dem $1/\sqrt{n}$-Gesetz resultierende Ungenauigkeit einbezieht, ist die geschätzte Wahrscheinlichkeit (der P-Wert) unter $0{,}0038 + 1/\sqrt{5000} = 0{,}0038 + 0{,}0141 = 0{,}0179$. Dieses zeigt eine deutliche Evidenz gegen unsere Annahme, „die Gruppenzugehörigkeit Koffein vs. kein Koffein hat keinen Einfluss auf die Fingertipprate".

5 Abschließende Bemerkungen

Das in diesem Artikel aufgeführte Beispiel soll das Potential von TinkerPlots bei der Durchführung von Randomisierungstests aufzeigen. Die Annahme (in diesem Fall „die Gruppenzugehörigkeit Koffein vs. kein Koffein hat keinen Einfluss auf die Fingertipprate") kann anschaulich und leicht in TinkerPlots simuliert werden. Die Zufallsmaschine bietet mit verschiedenen Bauteilen sowie mit der effektiven Befüllung dieser anhand von Copy&Paste gute Umsetzungsmöglichkeiten für Lernende, ohne dass Formelkenntnisse oder spezielle Softwarebefehle nötig wären. Entscheidend ist ferner, dass die Rerandomisierung per Knopfdruck ausgeführt werden kann und die – in unserem Fall 5000 – Messgrößen automatisch gesammelt werden können. Aus didaktischer Perspektive ist außerdem zu betonen, dass bei diesem Vorgehen und bei der Simulation die Visualisierung des Zufallsprozesses durch die Zufallsmaschine in TinkerPlots sehr gut dargestellt wird: Die Lernenden sehen die zufällige Zuordnung durch die Kugeln sowie das Etikettieren und Bekleben als Idee des Randomisierungstests und können so den Prozess nachvollziehen.

Durch vorbereitende „Hands-on-activities" wie die oben genannten wiederholten Kartenziehungen (siehe auch Arnold, Budgett & Pfannkuch, 2015) kann dieser Prozess vor der Simulation auf enaktiver Ebene erlebt werden.

Um die Lernenden strukturell und bei der Dokumentation der Durchführung ihres Randomisierungstests zu unterstützen und die Konzentration auf die konzeptionellen Aspekte beim Randomisierungstest zu legen, wurde nach dem Vorbild des Simulationsplans von Biehler & Maxara (2007) ein Randomisierungstestplan entwickelt (Biehler, Frischemeier & Podworny, 2015b). Dieser enthält die einzelnen Schritte, die beim Durchführen eines Randomisierungstests mit TinkerPlots zu bearbeiten sind und kann so die Arbeitsbelastung (vor allem den extraneous load) der Lernenden reduzieren und zum Strukturieren und Dokumentieren dienen.

Danksagung

Wir danken Joachim Engel sowie einem anonymen Reviewer für die vielen hilfreichen Anmerkungen und Kommentare bei der Überarbeitung dieses Artikels.

Anmerkungen

1 Siehe z. B. Hand, Daly, Lund, McConway & Ostrowski (1994, 40)

2 Die Ausprägungen des Merkmals „Gruppe" sind für die Durchführung des Randomisierungstests an dieser Stelle nicht notwendig, aber sie können anschaulich im weiteren Verlauf die Randomisierung deutlich machen (z. B. in Abb. 4), indem dann später die tatsächliche Ausprägung des Merkmals „Gruppe" mit der randomisierten Ausprägung des Merkmals „Etikett" verglichen werden kann.

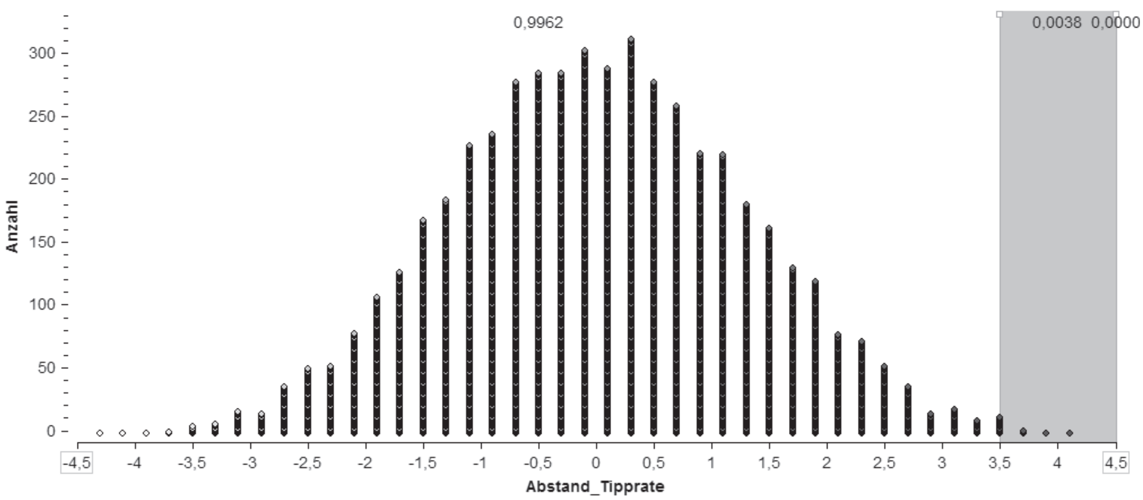

Abb. 8: Referenzgraphik zu den Messgrößen

Literatur

Arbeitskreis Stochastik der Gesellschaft für Didaktik der Mathematik. (2003). Empfehlung zu Zielen und zur Gestaltung des Stochastikunterrichts. *Stochastik in der Schule*, 23(3), 21–26.

Arnold, P., Budgett, S. & Pfannkuch, M. (2015). Experiment-to-causation inference: Understanding Causality in a Probabilistic Setting. In A. Zieffler & E. Fry (Eds.): *Reasoning about Uncertainty: Learning and Teaching Informal Inferential Reasoning.* Catalyst Press, pp. 95–128.

Biehler, R., Kombrink, K. & Schweynoch, S. (2003). MUFFINS – Statistik mit komplexen Datensätzen – Freizeitgestaltung und Mediennutzung von Jugendlichen. *Stochastik in der Schule*, 23(1), 11–25.

Biehler, R. (2007). TINKERPLOTS: Eine Software zur Förderung der Datenkompetenz in Primar- und früher Sekundarstufe. *Stochastik in der Schule*, 27(3), 34–42.

Biehler, R. & Maxara, C. (2007). Integration von stochastischer Simulation in den Stochastikunterricht mit Hilfe von Werkzeugsoftware. *Der Mathematikunterricht*, 53(3), 45–62.

Biehler, R., Ben-Zvi, D., Bakker, A. & Makar, K. (2013). Technology for Enhancing Statistical Reasoning at the School Level. In: K. Clements, A. Bishop, C. Keitel, J. Kilpatrick & F. Leung (Hrsg). *Third International Handbook of Mathematics Education.* New York: Springer 2013.

Biehler, R. & Frischemeier, D. (2015). „Verdienen Männer mehr als Frauen?" – Reale Daten im Stochastikunterricht mit der Software TinkerPlots erforschen. *Stochastik in der Schule*, 35(1), 7–18.

Biehler, R., Frischemeier, D. & Podworny, S. (2015a). Informelles Hypothesentesten mit Simulationsunterstützung in der Sekundarstufe I. *Praxis der Mathematik*, 66(6), 21–25.

Biehler, R.; Frischemeier, D. & Podworny, S. (2015b). Preservice Teachers' Reasoning about Uncertainty in the Context of Randomization Tests. In A. Zieffler & E. Fry (Eds.): *Reasoning about Uncertainty: Learning and Teaching Informal Inferential Reasoning (S. 129–162).* Catalyst Press.

Biehler, R., Frischemeier, D. & Podworny, S. (2016). Stochastische Simulationen mit TinkerPlots – Von einfachen Zufallsexperimenten zum informellen Hypothesentesten. *Stochastik in der Schule*, 36(1), 22–27.

Biehler, R., Kombrink, K. & Schweynoch, S. (2003). MUFFINS – Statistik mit komplexen Datensätzen – Freizeitgestaltung und Mediennutzung von Jugendlichen. *Stochastik in der Schule*, 23(1), 11–25.

Biehler, R. & Prömmel, A. (2013). Von ersten stochastischen Erfahrungen mit großen Zahlen bis zum $1/\sqrt{n}$-Gesetz – ein didaktisch orientiertes Stufenkonzept. *Stochastik in der Schule*, 33(2), 14–25.

Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1).

Erickson, T. (2006). Using simulation to learn about inference. In A. Rossman & B. Chance (Eds.), Working Cooperatively in Statistics Education. *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.

Hand, D. J., Daly, F., McConway, K., Lunn, D. & Ostrowski, E. (1993). *A handbook of small data sets.* Chapman and Hall: London.

KMK. (2004). Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss. München: Wolters Kluwer.

Konold, C. & Miller, C. (2011). TinkerPlots TM Version 2 [computer software]. Emeryville, CA: Key Curriculum Press.

Lock, R. H., Lock, P. F., Lock, K., Lock, E. & Lock, D. (2013). Statistics: *Unlocking the power of data.* Wiley Global Education.

Makar, K. and J. Confrey (2002). *Comparing two Distributions: Investigating Secondary Teachers' Statistical Thinking.* Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.

National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.

Rossman, A. (2008). „Reasoning about Informal Statistical Inference: A Statistician's View." *Statistics Education Research Journal*, 7(2), 5–19.

Vogel, M. (2009). Experimentieren mit Papierfröschen, *PM – Praxis der Mathematik in der Schule*, 51(2), 22–30.

Watson, J. M. & Moritz, J. B. (1999). The beginnings of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37, 145–168.

Watson, J. M. (2013). Resampling with TinkerPlots. *Teaching Statistics*, 35(1), 32–36

Anschrift der Verfasser

Rolf Biehler
Institut für Mathematik
Universität Paderborn
Warburger Straße 100
33098 Paderborn
biehler@math.upb.de

Daniel Frischemeier
Institut für Mathematik
Universität Paderborn
Warburger Straße 100
33098 Paderborn
dafr@math.upb.de