

Daniel FRISCHEMEIER, Paderborn

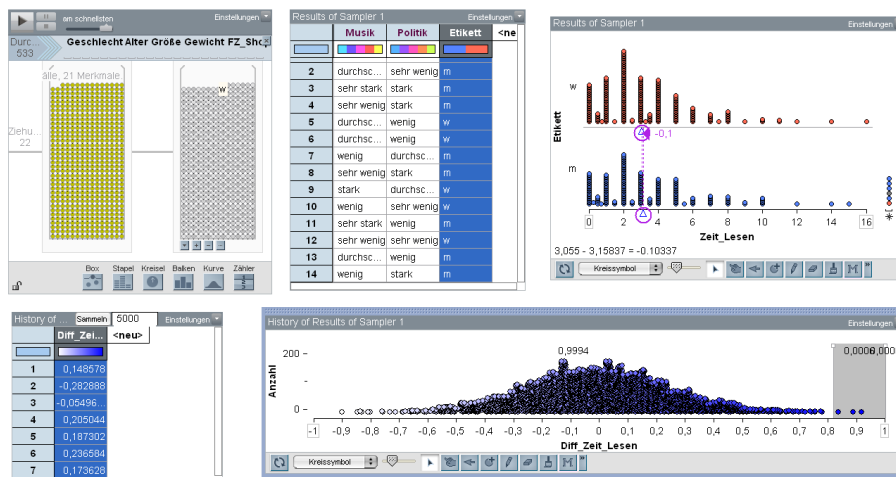
Verteilungen vergleichen mit TinkerPlots – und darüber hinaus weitere Schlussfolgerungen aus Daten generieren

Vergleiche von Verteilungen zweier quantitativer Merkmale (kurz: Verteilungsvergleiche) sind ein fundamentaler Bestandteil der deskriptiven Statistik. Wünschenswerterweise sind diese Verteilungsvergleiche in einem Datenanalysezyklus (z.B. im PPDAC-Zyklus nach Wild & Pfannkuch 1999) eingebettet, so dass diese anhand von selbsterhobenen realen Daten und selbst generierten Fragestellungen und Hypothesen durchgeführt werden. Nun liegt es nahe in einem nächsten Schritt zu versuchen, weitere Schlüsse aus den Daten zu ziehen, z.B. könnte man die Frage aufwerfen, inwieweit sich die Beobachtungen (Unterschiede zwischen den Verteilungen) verallgemeinern lassen oder ob sie nur zufällig (aufgrund der Stichprobe) entstanden sind. In diesem Artikel möchten wir eine Möglichkeit vorstellen über Verteilungsvergleiche dieser Art mithilfe von Randomisierungstests hinauszugehen und weitere Schlüsse aus den Daten zu ziehen.

1. Randomisierungstests

Neben den erhobenen Daten anhand eines selbstkonstruierten Fragebogens im Rahmen einer Umfrage unter Studienanfängern an der Universität Paderborn haben wir im Seminar (welches in Frischemeier & Biehler (2012) näher beschrieben ist) auch mit dem Muffins-Datensatz (Biehler et al., 2003) gearbeitet. Dieser Datensatz enthält über 50 Variablen zum Freizeitverhalten von 538 Schülerinnen und Schülern aus Nordrhein-Westfalen. Exemplarisch für einen Verteilungsvergleich auf Grundlage dieser Daten ist beispielsweise die Fragestellung: „Wie unterscheidet sich der Umfang der Computernutzung zwischen Jungen und Mädchen?“ (Biehler et al., 2003). Hier schließen sich, wie im obigen Abschnitt schon angedeutet, weitere Fragen an - beispielsweise: Gibt es eine Möglichkeit die Aussage zu verallgemeinern? Theoretisch könnte ein möglicher Unterschied zwischen beiden Gruppen rein zufällig (aufgrund unserer Stichprobe) entstanden sein. Wie kann man Lernenden ein Hilfsmittel oder Anhaltspunkte geben, um weitere Schlussfolgerungen aus vorliegenden Daten wie in diesem Beispiel zu generieren? Eine Alternative und ein mögliches Instrument um oben aufgeworfene Fragen zumindest qualitativ beantworten zu können, sind Randomisierungstests (Rossman, 2008, 10). Cobb (2007, 12) schlägt vor, dass Studierenden in statistischen Einführungskursen durch Randomisierungstests die Möglichkeit gegeben werden sollte die Kernideen der Inferenz zu verstehen. Dabei verweist Cobb auf drei wichtige Aspekte und die Schrittfolge beim Durchführen eines solchen Tests: „1.

Randomize data production, 2. Repeat by simulation to see what's typical (...and what's not), 3. Reject any model that puts your data in its tail". An dieser Stelle müssen wir für die bei uns verwendeten Muffins-Daten eine Einschränkung machen, denn es handelt sich hier um keine Zufallsstichprobe bzw. randomisierte Stichprobe (wie bei Cobb in 1. gefordert), sondern um „observational data“. Argumente für die Durchführung eines Randomisierungstests auch unter diesen Bedingungen finden sich u.a. in Zieffler et al. (2011). Ebenso sind sich sowohl Rossman (2008) als auch Cobb (2007) einig, dass einerseits die Umsetzung ohne eine Unterstützung durch adäquate Software schwierig zu sein scheint, andererseits die Durchführung eines solchen Tests auch nicht zu software-technisch sein soll. Die Software TinkerPlots (Konold & Miller, 2011) mit ihrer visualisierten Zufallsmaschine scheint für die Durchführung in besonderem Maße geeignet zu sein. Die folgende Abbildung zeigt die Arbeitsoberfläche der Software TinkerPlots bei der Durchführung eines Randomisierungstests.



Das Vorgehen und die Durchführung eines solchen Tests mit TinkerPlots können in Frischemeier & Biehler (2013) ausführlich nachgelesen werden.

2. Eine Lernumgebung zu Randomisierungstests

Eine speziell für das Erlernen von Randomisierungstests konstruierte Lernumgebung sah neben der Einführung in die Zufallsmaschine in TinkerPlots und der Simulation einiger Zufallsexperimente das Durchführen eines Randomisierungstests anhand der Muffins-Aufgabe („Gibt es wirklich einen geschlechtsspezifischen Unterschied zwischen Jungen und Mädchen hinsichtlich ihrer Lesegewohnheiten (Merkmal: Zeit_Lesen) oder ist dieser Unterschied zufällig – aufgrund der Wahl der Stichprobe – entstanden?) vor. Als Unterstützung bekamen die Lernenden einerseits einen 6-schrittigen Randomisierungstest-Plan, der die Schrittfolge des Tests vorstrukturierte und die Schritte jeweils erläuterte, sowie ein Handout mit

„Faustregeln“ zur Beurteilung verschiedener P-Werte. Der Randomisierungstest-Plan konstruiert nach dem Vorbild eines Simulationsplans (Maxara & Biehler, 2007) sollte den Lernende eine strukturierende Unterstützung bieten und den „extraneous load“ reduzieren.

3. Eine explorative Fallstudie

Im Folgenden wollen wir eine explorative Fallstudie zu Randomisierungstests mit TinkerPlots vorstellen, die wir zum Ende des Seminar „Statistisch denken und forschen lernen“ durchgeführt haben. Im besagten Seminar mussten die 22 Teilnehmer gemeinsam zu zweit eine statistische Projektarbeit als „Gesellenstück“ für einen erfolgreichen Abschluss des Seminars anfertigen. Diese Projektarbeit enthielt neben einigen Verteilungsvergleichen auch einen Randomisierungstest. Die Randomisierungstests (die von den einzelnen Paaren zu jeweils unterschiedlichen Fragestellungen durchgeführt wurden) in schriftlicher Form haben wir hinsichtlich von drei explorativen Forschungsfragen analysiert, um die Wirksamkeit unserer Lernumgebung einschätzen zu können: 1. Wie führen die Teilnehmer nach dieser kurzen Einführung einen Randomisierungstest in TinkerPlots durch? 2. Sind sie nach dieser kurzen Einführung in der Lage die sechs Schritte (Randomisierungstest-Plan) zu vollziehen? 3. An welchen Stellen treten Probleme (welche?) auf? Bei der Auswertung haben wir zwischen zwei Ebenen unterschieden: einer globalen und einer lokalen Ebene. Wir werden im Folgenden die wichtigsten Erkenntnisse der globalen Ebene beschreiben. Hier haben wir eine Häufigkeitsanalyse (Mayring, 2010), unter der Fragestellung, inwieweit die einzelnen durch den Randomisierungstest-Plan vorgegebenen sechs Schritte nach dem Abschluss der Lernumgebung sicher/ korrekt durchlaufen werden, durchgeführt.

Schritt erfolgreich durchgeführt	Anzahl der Teams (von 11 insgesamt)	%
Schritt 1 (Beobachtung)	11	100,00
Schritt 2 (Nullhypothese)	8	72,73
Schritt 3 (Simulation der Nullhypothese)	10	90,91
Schritt 4 (Teststatistik korrekt erzeugt)	10	90,91
Schritt 5 (p-Wert korrekt abgelesen)	5	45,45
Schritt 6 (korrekte Folgerungen aus p-Wert)	5	45,45

Beim Blick auf die Tabelle fällt auf, dass prozedurale Aktivitäten (wie z.B. die Durchführung des Tests mit der Software) gelingen, eher konzeptionelle Aktivitäten (wie das korrekte Aufstellen der Nullhypothese oder das Ablesen und Beurteilen des P-Wertes) aber Schwierigkeiten bereiten. Der lo-

kale Blick auf die Daten (Ausschnitte finden sich in Frischemeier & Biehler, 2013) zeigt Schwierigkeiten beim Beurteilen eines P-Wertes, insbesondere wenn dieser größer als 0,1 ist. Alle Teilnehmer, die einen Randomisierungstest durchführten und einen P-Wert größer als 0,1 erhielten, haben falsche Schlussfolgerungen (im Sinne von: „die Nullhypothese ist richtig“ / „die Nullhypothese kann bestätigt werden“) daraus gezogen.

4. Fazit & Ausblick

Zur Fallstudie in diesem Artikel bleibt zu sagen, dass es sich um eine explorative Studie handelte und bei der Auswahl der Probanden beispielsweise keine Zufallsauswahl vorlag, so dass verallgemeinernde Aussagen nur schwer möglich sind. Außerdem hat es aufgrund von Zeitmangel nur eine mehr oder weniger knappe Einführung in das Feld der Randomisierungstests gegeben. Hier muss in Zukunft an den kritischen Stellen (Aufstellen der Nullhypothese, Ablesen und Beurteilen von P-Werten) mehr Zeit investiert und die Lernumgebung überarbeitet werden.

Literatur

- 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. & Maxara, C. (2007). Integration von stochastischer Simulation in den Stochastikunterricht mit Hilfe von Werkzeugsoftware. *Der Mathematikunterricht* 53 (3): 45-62.
- Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1).
- Frischemeier, D. & Biehler, R. (2012). Statistisch denken und forschen lernen mit der Software TinkerPlots. In Kleine, M. und Ludwig, M. (Eds.): *Beiträge zum Mathematikunterricht 2012*, WTM: Münster.
- Frischemeier, D. & Biehler, R. (2013) Design and exploratory evaluation of a learning trajectory leading to do randomization tests facilitated by TinkerPlots. *Proceedings of CERME-8* (submitted).
- Konold, C. & Miller, C. (2011). *TinkerPlots TM Version 2* [computer software]. Emeryville, CA: Key Curriculum Press.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 11. Auflage. Weinheim: Beltz.
- Rossman, A. (2008). "Reasoning about Informal Statistical Inference: A Statistician's View." *Statistics Education Research Journal* 7(2): 5-19.
- Wild, C. J. and M. Pfannkuch (1999). "Statistical Thinking in Empirical Enquiry." *International Statistical Review* 67(3): 223-265
- Zieffler, A., Haring, J., & Long, J. (2011). *Comparing groups: Randomization and bootstrap methods using R*. New York: Wiley.