

Kompetenzen von Lehramtsstudierenden beim Vergleich von Verteilungen unter Nutzung der Software TinkerPlots

DANIEL FRISCHEMEIER, PADERBORN

Zusammenfassung: Der Vergleich von Verteilungen eines numerischen Merkmals ist eine fundamentale Aktivität in der deskriptiven Statistik. Um dieser Aktivität auch bei realen, multivariaten Datensätzen nachgehen zu können, benötigt man eine geeignete Datenanalyse-Software, um die Daten je nach Fragestellung drehen und wenden zu können. In diesem Artikel wird untersucht, wie Studierende des Lehramts Mathematik an Grund-, Haupt-, Real- und Gesamtschulen mithilfe der Datenanalyse-Software TinkerPlots Verteilungen in realen und multivariaten Datensätzen vergleichen.

Abstract: Group comparisons are a fundamental activity in descriptive statistics. In order to be able to compare groups also in real and multivariate data sets, the use of an adequate digital data analysis tool is inevitably to explore the data in regard to meaningful statistical questions. The study reported in this article examines how preservice teachers for mathematics at primary and secondary schools compare distributions in real and multivariate data sets using the data analysis software TinkerPlots.

1. Einleitung

Durch die Etablierung der Bildungsstandards in der Primarstufe (Hasemann & Mirwald, 2012) und in der Sekundarstufe I (Blum, Drüke-Noe, Hartung & Köller, 2006) hat die Stochastik im Mathematikunterricht in den letzten Jahren ein stärkeres Gewicht und eine größere Bedeutung bekommen. Im Bereich der Leitidee Daten werden dabei u. a. Aktivitäten wie das Durchlaufen eines Datenanalysezyklus, das Interpretieren der Ergebnisse der Datenanalyse sowie der Einsatz adäquater (Datenanalyse-)Software vorgesehen. Die Umsetzung dieser Standards im Unterricht stellt eine Herausforderung nicht nur für Lehrer¹, sondern insbesondere auch für die Hochschule dar, die mit der Ausbildung der Lehrer beauftragt ist. Die Hochschule muss neben den didaktischen Komponenten („pedagogical knowledge“) zunächst erst einmal die fachlichen („content knowledge“) und die technologischen („technological knowledge“) Wissenskomponenten fördern (Mishra & Koehler, 2006).² Für Zweiteres (fachliche Wissenskomponenten) existieren nationale (Arbeitskreis Stochastik der Gesellschaft für Didaktik der Mathematik, 2012) und internationale Empfehlungen (Pfanckuch & Ben-Zvi, 2011) zur Ausbildung des fachlichen Wissens von

Lehrkräften. So fordern Pfanckuch and Ben-Zvi (2011) für die Ausbildung von Lehrkräften in Stochastik den Ausbau der fachlichen Kompetenzen als erstes vorrangiges Ziel im Hinblick auf die veränderten Anforderungen im Bereich der Stochastik.

The first goal is to develop and improve teachers' understanding of statistics, since it is generally acknowledged that they lack statistical knowledge, good statistical thinking and points of view that are now required by new curricula [...]. (Pfanckuch & Ben-Zvi 2011, S. 324)

Neben dem selbstständigen Durchführen einer Datenerhebung fordern Pfanckuch und Ben-Zvi (2011) auch das Anwenden vielfältiger statistischer Konzepte und Darstellungen beim Prozess der Datenanalyse. Ebenso werden ein kompetenter Umgang mit Datenanalyse-Software sowie die Interpretation grafischer Darstellungen gefordert. Um diese Kompetenzen in der universitären Lehrerbildung entsprechend auszubauen, existieren besonders in der internationalen Stochastikdidaktik – siehe vor allem im Sammelband von Garfield and Ben-Zvi (2008) – viele Anregungen und Umsetzungsideen für eine Lehrveranstaltung zur Statistik für Lehramtsstudierende. Dabei wird vor allem auf das Durchlaufen eines kompletten Datenanalysezyklus wie den sogenannten *PPDAC-Zyklus* Wert gelegt, der nach Wild und Pfanckuch (1999) die Komponenten Problem (Generieren einer statistischen Fragestellung auf Grundlage eines statistischen Problems), Plan (Planen einer Datenerhebung und Konstruktion eines geeigneten Datenerhebungsinstruments), Data (Erhebung der Daten), Analysis (Exploration und Analyse der Daten) sowie Conclusions (Interpretation der Daten und Beantwortung der vorausgegangenen statistischen Fragestellung) umfasst. Darüber hinaus wird das Arbeiten mit realen Daten (Garfield & Ben-Zvi, 2008) und der Einsatz adäquater Software (Biehler, Ben-Zvi, Bakker & Makar, 2013) als sehr bedeutsam hervorgehoben.

Insbesondere im Bereich der Datenanalyse erscheinen Aktivitäten zum Vergleich von Verteilungen eines numerischen Merkmals als sehr substanziell (Konold, Pollatsek, Well & Gagnon, 1997), weil sich Lernende dort mit vielen fundamentalen Ideen der Stochastik (nach Burrill & Biehler, 2011, wie Daten, Repräsentation, Verteilung, Variabilität, etc.) auseinandersetzen können. Verteilungsvergleiche sind

Aktivitäten, die an Fragestellungen wie etwa „Inwiefern unterscheiden sich die Schülerinnen und Schüler hinsichtlich ihres Fernsehkonsums?“ (siehe z. B. Biehler, Kombrink & Schweynoch, 2003) anschließen. Verteilungsvergleiche werden auch normativ in Statistikcurricula gefordert und sind in der Datenanalyse im Besonderen von fundamentaler Bedeutung (Watson & Moritz, 1999; Makar & Confrey, 2002). Laut Konold and Higgins (2003) sind Verteilungsvergleiche besonders dann lohnenswert, wenn man sie in realen und multivariaten Datensätzen durchführt und eine Software verwendet, die ein Drehen und Wenden der Daten nach selbstgewählten Fragestellungen zulässt. In Kombination mit dem Einsatz adäquater Software laden Verteilungsvergleiche dazu ein, eigene Vergleichshypothesen in reichhaltigen Datensätzen aufzustellen und diese mittels Software zu explorieren und darüber hinaus weitere Hypothesen (z. B. für Gründe der vorliegenden Unterschiede) zu generieren. Weitere Details zu verschiedenen Formen von Verteilungsvergleichen finden sich in Biehler, Frischemeier, Reading, and Shaughnessy (2018, S. 163 ff.).

Die Software *TinkerPlots* (Konold & Miller, 2011)³, die als Lernsoftware auf Basis internationaler Erkenntnisse der Stochastikdidaktik entwickelt wurde, ist eine Lernsoftware, die für die Förderung statistischer und probabilistischer Denkweisen von Schülern ab Klasse 4 vorgesehen ist und auch in der Lehrerausbildung gewinnbringend eingesetzt werden kann, um Datenanalyse zu betreiben (Podworny, Frischemeier & Biehler, 2017). Für die Benutzung der Software *TinkerPlots* wird keine Formelkenntnis und daher nur eine geringe Lernzeit benötigt (Konold, 2006, 2007). Sie zeichnet sich dadurch aus, dass an das Arbeiten und an die Operationen mit Datenkarten (Trennen, Ordnen, Stapeln) angeknüpft wird (Harradine & Konold, 2006) und Graphiken im weiteren Sinne nur basierend auf den Grundoperationen „Stapeln“, „Ordnen“ und „Trennen“ erstellt werden können. *TinkerPlots* kann als Lernsoftware für Schüler ab Jahrgangsstufe 4, als Datenanalysesoftware für Studierende in der Lehrerbildung und außerdem vom Lehrer später als Demonstrationsmedium im Unterricht eingesetzt werden (Biehler et al., 2013). Die Forschung rund um den Einsatz der Software *TinkerPlots* und der Ausbau und die Entwicklung einer Datenanalysekompetenz mit dieser Software stehen bei einigen internationalen Forschergruppen im Bereich der Stochastikdidaktik im Fokus: in den USA (u. a. Rubin, Hammerman & Konold, 2006) in Australien (u. a. Fitzallen & Watson, 2014) und in Israel (u. a. Gil & Ben-Zvi, 2014). Allerdings gibt es derzeit nach eigenen Recherchen und bestem Wissen zum Stand der Dinge international bisher nur sehr wenige Stu-

dien (national keine) zum Einsatz der Software *TinkerPlots* von Lehramtsstudierenden beim Bearbeiten von Datenanalyseaufgaben. Auch das Potential der Software *TinkerPlots* beim Verteilungsvergleich sowie entsprechende Verteilungsvergleichskompetenzen wurden noch nicht erforscht. Aus diesem Grund war das wesentliche Ziel der diesem Artikel zugrundeliegenden Studien, zu untersuchen wie Lehramtsstudierende der Mathematik an Grund-, Haupt-, Real- und Gesamtschulen (im Weiteren verkürzt: GHRGe) Verteilungen unter Nutzung der Software *TinkerPlots* vergleichen. Die entsprechende Forschungsfrage dazu lautet: *Wie vergleichen Studierende des Lehramts Mathematik an Grund-, Haupt-, Real- und Gesamtschulen zwei gegebene Verteilungen eines numerischen Merkmals mit der Software TinkerPlots?* Dabei stehen die beiden Ziele *Analyse der statistischen Kompetenzen* (Wie vergleichen Lehramtsstudierende Verteilungen?) und die *Analyse der Softwarekompetenzen* (Wie nutzen Lehramtsstudierende *TinkerPlots*, um Verteilungen zu vergleichen?) der Teilnehmer gleichberechtigt nebeneinander. Die in diesem Artikel ausgeführten Analysen und Untersuchungen sowie die Ergebnisse der Untersuchungen beruhen zum Teil auf einem Dissertationsprojekt (Frischemeier, 2017), welches im Jahr 2016 abgeschlossen wurde.

2. Verteilungsvergleiche

Im Folgenden wird zunächst ein Überblick über tragfähige Verteilungsvergleichselemente (2.1) gegeben und es werden dann bereits existente empirische Studien reflektiert, die untersucht haben, wie Lernende Aufgaben zum Verteilungsvergleich bearbeiten (2.2). Das Ziel dieses Abschnitts ist es, deduktiv ein Schema zur Evaluation von Vorgehensweisen beim Verteilungsvergleich zu konzipieren, um dieses für die Auswertung der Daten einer Fallstudie anwenden zu können.

2.1 Tragfähige Verteilungsvergleichselemente

Es stellt sich die Frage, welche Elemente beim Verteilungsvergleich als tragfähig angesehen werden können. Rossman, Chance, and Lock (2001, S. 48) sehen *Zentrum*, *Streuung*, *Form*, *Teilgruppen* und *Ausreißer* als fundamentale Bestandteile einer Verteilung an. Diese Komponenten lassen sich auch auf den Vergleich von Verteilungen übertragen. Pfannkuch (2007) hat darüber hinaus Elemente definiert, die sich beim Vergleichen von Boxplots herausarbeiten lassen wie *Signal*, *Shift*, *Summary* und *Spread*. Dabei verwendet Pfannkuch das Verteilungsvergleichselement *Signal* wenn Unterschiede zwischen den mittleren 50 % von Verteilungen herausgearbei-

D. Frischemeier

tet werden. Pfannkuch (2007) spricht von dem Verteilungsvergleichselement *Shift*, wenn eine (z. B. additive oder multiplikative) Verschiebung zwischen Verteilungen identifiziert wird. Beim Verteilungsvergleichselement *Summary* werden Verteilungen anhand der 5-Zahlen-Zusammenfassung (Minimum, 1. Quartil, Median, 3. Quartil und Maximum) der jeweiligen Verteilung, beim Verteilungsvergleichselement *Spread* anhand der Streuung der Verteilungen des jeweiligen numerischen Merkmals verglichen. Biehler (2001, S. 110, 2007a und 2007b) schlägt normativ vor, auch die Form von Verteilungen beim Vergleich miteinzubeziehen. Außerdem nennt er individuelle Vergleichsansätze wie p-basierte und q-basierte Vergleiche (Biehler, 2001, S. 110) als weitere Möglichkeiten zum Vergleich zweier Verteilungen. Ein p-basierter Vergleich liegt nach Biehler (2001, S. 110) vor...

...wenn für ein x aus dem Wertebereich die relativen Häufigkeiten $h(V \leq x)$ und $h(W \leq x)$ verglichen werden, bzw. dasselbe mit \geq . Man gibt eine Grenze vor, z. B. 10 Stunden und vergleicht den Anteil derjenigen, die größer gleich 10 Stunden lesen in beiden Gruppen. (Biehler, 2001, S. 110)

Ein q-basierter Vergleich wird von Biehler (2001, S. 110) so beschrieben:

Einen Vergleich nenne ich „q-basiert“, wenn für einen Anteil p zwischen 0 und 1 die zusammenpassenden Quantile der beiden Variablen V und W , $q_V(p)$ mit $q_W(p)$, verglichen werden (mit $q(p)$ ist das Quantil zu p gemeint). Bei $p = 0,5$ bedeutet das einfach einen Vergleich der Mediane. (Biehler, 2001, S. 110)

Als tragfähige Verteilungsvergleichselemente lassen sich somit auf Grundlage der in diesem Abschnitt zitierten Literatur die folgenden Elemente (siehe auch Frischemeier, 2017, S. 33) herauskristallisieren: *Zentrum, Streuung, Verschiebung, Form, p-basiert und q-basiert*.

2.2 Empirische Studien zum Verteilungsvergleich

Die empirischen Studien zum Verteilungsvergleich reichen von Untersuchungen bei Grundschulern (Watson & Moritz, 1999) über Schüler der Sekundarstufe (Pfannkuch, Budgett, Parsonage & Horing, 2004; Pfannkuch, 2007) bis hin zu Untersuchungen mit Universitätsstudierenden (Makar & Confrey, 2014). Insgesamt lassen sich einige Unzulänglichkeiten beim Vergleichen von Verteilungen entdecken, so z. B. dass Ergebnisse nicht interpretiert (vgl. Biehler, 1997; Pfannkuch, 2007) werden oder dass beim Vergleichen eine Beschränkung auf ein oder wenige Verteilungsvergleichselemente, wie z. B. das arithmetische Mittel, stattfindet (vgl. Pfannkuch, 2007; Frischemeier & Biehler, 2011).

Pfannkuch et al. (2004) sowie Pfannkuch (2007) haben die Verteilungsvergleichskompetenz von Schülern im Alter von 14 und 15 Jahren untersucht. In der empirischen Untersuchung von Pfannkuch et al. (2004) bekamen die Probanden die Verteilungen von Temperaturdaten aus den neuseeländischen Städten Napier und Wellington vorgelegt und sollten diese vergleichen (Pfannkuch et al. 2004, S. 3). In einer strukturierenden Analyse wurden zunächst verschiedene Aussagetypen beim Vergleich zweier Boxplots herausgearbeitet: *conclusion* (z. B. „In Napier ist es wärmer als in Wellington“), *comparing equivalent summary statistics* (z. B. „Napier hat einen höheren Median als Wellington“), *comparing non-equivalent summary statistics* (z. B. „Der Median von Napier ist größer als das 3. Quartil von Wellington“), *comparing variability* (z. B. Die Temperaturen in Napier haben eine größere Spannweite verglichen mit denen in Wellington“) und *comparing distributions* (z. B. „Der Boxplot für Wellington ist weiter links als der für Napier“). In einem darauffolgenden Analyseschritt in Bezug auf die Qualität der einzelnen Vergleichsaussagen wurden die Aussagen in den einzelnen Bereichen (*conclusions* etc.) nach dem Vorbild der SOLO Taxonomie (Biggs & Collis, 1982) in *non-response, prestructural, unistructural, multistructural* und *relational* eingestuft. Ein wesentliches Resultat der Studie ist, dass die Schüler die Verteilungen vor allem anhand von *summary statistics*, aber nur sehr selten anhand von *variability* und *comparing distributions* vergleichen. Außerdem ist auffällig, dass in den Kategorien *conclusions* und *comparing equivalent summary statistics* noch verhältnismäßig viele Aussagen im Bereich *multistructural* oder *relational* kodiert wurden. Dieses nimmt hin zu den filigraneren Vergleichsmethoden wie *comparing variability* oder *comparing distributions* ab. Es zeigt, dass Lernende gerade bei Vergleichen, die sich nicht nur auf Mittelwerte oder auf *summary statistics* beschränken, Probleme haben, die Verteilungen anhand der hier definierten Elemente *comparing variability* und *comparing distributions* zu vergleichen und die herausgearbeiteten Unterschiede in Zusammenhang bringen.

In einer Folgestudie anknüpfend an Pfannkuch et al. (2004) hat Pfannkuch (2007) die Verteilungsvergleichskompetenzen von 15-jährigen Schülern untersucht. Die Probanden haben dabei Verteilungen des Merkmals „Anzahl von Kurzmitteilungen“ in Form von Boxplots vorgelegt bekommen und wurden dann aufgefordert, die Verteilungen zu vergleichen. Die Mitschriften der Probanden wurden eingesammelt und mit qualitativen Methoden ausgewertet. Ähnlich wie bei Pfannkuch et al. (2004) ist die Auswertung der Daten zweistufig vorgenommen worden. In einer ersten Stufe wurden die Aussagen strukturiert und in

einer zweiten Stufe hinsichtlich der Qualität bewertet. Unter anderem – siehe Pfannkuch (2007, S. 159) für Details – wurden die Vergleichsaussagen in die Kategorien *summary*, *signal*, *shift* und *spread*, die in Abschnitt 2.1 schon ausgeführt worden sind, eingestuft. Auf einer zweiten, skalierenden Ebene wurden die Qualitäten der Vergleichsaussagen in die Stufen *point decoder*, *shape comparison describer*, *shape comparison decoder* und *shape comparison assessor* eingestuft. Während auf der Ebene *point decoder* Lernende Kennzahlen lediglich identifizieren aber nicht vergleichen, findet auf den Ebenen *shape comparison describer* und *shape comparison decoder* ein Vergleich statt. Die Unterschiede zwischen den Verteilungen werden dabei aber lediglich beschrieben (z. B. das arithmetische Mittel ist in Gruppe A höher als in Gruppe B) und nicht interpretiert. *Shape comparison assessor* ist die höchste Kompetenzstufe bei Pfannkuch (2007). Auf dieser Stufe werden Unterschiede beim Verteilungsvergleich nicht nur beschrieben, sondern auch im Kontext interpretiert. Ein wesentliches Ergebnis der empirischen Studie von Pfannkuch (2007) ist, dass die Lernenden, die an der Studie teilgenommen haben, herausgearbeitete Unterschiede zwischen Verteilungen selten interpretieren, sondern eher auf der beschreibenden Ebene bleiben.

Es existieren bislang keine Studien und Schemata, die bewerten, wie kompetent Lernende digitale Werkzeuge im Datenanalyseprozess einsetzen. Maxara (2009, 2014) hat untersucht, inwieweit Studierende die Software Fathom beim Simulieren von Zufallsexperimenten kompetent nutzen können und hat ein Modell für Softwarekompetenz aufgestellt, welches auch für die Einschätzung einer TinkerPlots-Softwarekompetenz bei der Datenanalyse nützlich sein könnte. Maxara (2014, S. 327) selbst erläutert, dass sich

die [...] vier Kompetenzbereiche zur Fathom-spezifischen Simulationskompetenz [...] prinzipiell auch auf andere Software und Anwendungsbereiche übertragen [lassen]. Die allgemeinen Fathom-Kompetenzen, die Formelkompetenz sowie die strategischen Kompetenzen lassen sich ohne weiteres auf eine andere Werkzeugsoftware wie z. B. Excel übertragen. Dann würden beispielsweise die allgemeinen Excel-Kompetenzen die technischen Kompetenzen im Umgang mit den grundlegenden Objekten und Werkzeugen in Excel umfassen usw. [...] (Maxara, 2014, S. 327)

Im Hinblick auf die Forschungsfragestellung „Wie vergleichen Studierende des Lehramts Mathematik an Grund-, Haupt-, Real- und Gesamtschulen zwei gegebene Verteilungen eines numerischen Merkmals mit der Software TinkerPlots?“ lassen sich zwei Dimensionen, die es beim Verteilungsvergleich mit Software zu beachten gibt, ableiten:

Die Dimension *TinkerPlots-Kompetenz* zur Analyse der Softwarekompetenzen (siehe Ziele aus Abschnitt 1) beleuchtet – in Anlehnung an die Vorarbeiten von Maxara (2009; 2014) zur Untersuchung der kompetenten Nutzung der Software Fathom beim Simulieren von Zufallsexperimenten mit Fathom – wie kompetent die Software TinkerPlots im Verteilungsvergleichsprozess eingesetzt wird.

Die Dimension *Verteilungsvergleichskompetenz* zur Analyse der statistischen Kompetenzen (siehe Ziele aus Abschnitt 1) bildet die Teilkompetenzen ab, die nötig sind, um Unterschiede zwischen Verteilungen in Anlehnung an Biehler (2001), Pfannkuch et al. (2004), Pfannkuch (2007) und Biehler (2007a, 2007b) adäquat herauszuarbeiten. Fragen, die im Rahmen dieser Kompetenz im Vordergrund stehen, sind: „Welche Aspekte nutzen die Studierenden beim Vergleichen von Verteilungen?“ und „Welche Qualität haben ihre Vergleichsaussagen?“

2.3 Deduktiv abgeleitete Schemata zum Verteilungsvergleich

Im Folgenden werden zwei Analyseinstrumente zur Auswertung vorgestellt, die auf der Grundlage der Abschnitte 2.1 und 2.2 deduktiv abgeleitet werden: Das Kategoriensystem zur *TinkerPlots-Kompetenz* und das Kategoriensystem zur *Verteilungsvergleichskompetenz*. Die Bildung der Kategorien wurde durch Kombination aus deduktivem und induktivem Vorgehen (siehe Kuckartz, 2012) durchgeführt. Auf Grundlage der theoretischen Vorarbeiten in 2.1 und 2.2 wurden so zunächst zu den zwei Dimensionen *TinkerPlots-Kompetenz* und *Verteilungsvergleichskompetenz* deduktive Kategorien gebildet, die beim Durchlauf durch die Daten einer Fallstudie (siehe Abschnitt 3, Transkripte & Videos) dann induktiv angereichert und verfeinert wurden.

2.3.1 Kategoriensystem *TinkerPlots-Kompetenz*

Das Kategoriensystem *TinkerPlots-Kompetenz* soll untersuchen, inwieweit Lernende die Software TinkerPlots kompetent für ihre Verteilungsvergleichsvorhaben nutzen können. Auf dieser Grundlage wurde das Kategoriensystem *TinkerPlots-Kompetenz* deduktiv auf Basis der Vorarbeiten zur Fathom-Kompetenz nach Maxara (2009, 2014) entwickelt und die Ausprägungen hohe/mittlere/niedrige *TinkerPlots-Kompetenz* etabliert. Insbesondere wird in der hier vorliegenden Studie die *TinkerPlots-Kompetenz* so operationalisiert, dass untersucht wird, inwieweit geplante statistische Aktivitäten tatsächlich mit TinkerPlots umgesetzt werden können. So ist eine hohe *TinkerPlots-Kompetenz* gegeben, wenn in der Phase der *statistischen Aktivität* (bevor die Software genutzt wird) eine Aktion mit der Software formuliert wird

D. Frischemeier

und diese mit dem Werkzeug fehlerfrei umgesetzt wird. Als Beispiel das Vorgehen von Laura und Ricarda (Namen geändert) aus der Fallstudie:

Laura: „Dann lass uns erst mal einfach mit dem Mittelwert anfangen“

Laura und Ricarda klicken auf „arithmetisches Mittel“ und „Wert anzeigen“.

Laura formuliert hier ihre Intention, einen Mittelwert zu berechnen. Dieses wird daraufhin in TinkerPlots korrekt umgesetzt (siehe TinkerPlots Graph in Abb. 1), was in diesem Fall auf eine hohe *TinkerPlots-Kompetenz* schließen lässt. Eine mittlere *TinkerPlots-Kompetenz* liegt vor, wenn eine konkrete Intention zur Erstellung einer Graphik oder zum Berechnen eines Wertes in TinkerPlots zum Verteilungsvergleich formuliert wird, dieses aber nur mit einem trial-and-error-approach realisiert werden kann. Ein Beispiel für einen „trial-and-error-approach“ ist beim Vorgehen von Hilde und Iris (ebenfalls aus der Fallstudie, Namen geändert) gegeben: Hilde und Iris verfolgen die Intention das arithmetische Mittel zu berechnen, wissen aber nicht, welchen Schalter sie in TinkerPlots dazu benutzen müssen und finden diesen nur durch Probieren. Schließlich wird von einer niedrigen *TinkerPlots-Kompetenz* gesprochen, wenn eine konkrete Intention zum Verteilungsvergleich vorliegt, diese aber nicht mit TinkerPlots umgesetzt werden kann. Ein Beispiel aus der Fallstudie ist unter anderem bei Laura und Ricarda zu finden: Laura und Ricarda möchten einen Boxplot in TinkerPlots erstellen. Nach einiger Zeit sagt Laura „Ich weiß nicht wie es geht“ und sie brechen ihr Vorhaben ab, die Verteilungen anhand von Boxplots zu vergleichen. Dass Laura und Ricarda hier lokal ihre geplanten und artikulierten Vergleichsaktivitäten mit TinkerPlots nicht umsetzen können, ist an dieser Stelle ein Indikator für eine niedrige *TinkerPlots-Kompetenz*.

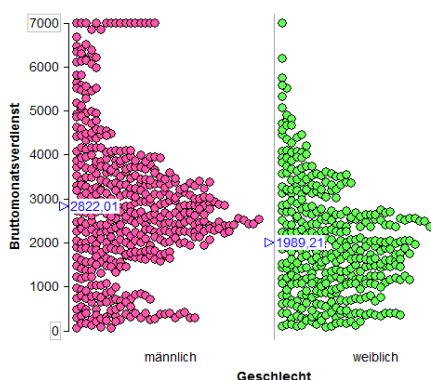


Abb. 1: TinkerPlots Graph von Laura und Ricarda

In der Tabelle 5 im Anhang sind zusammengefasst die einzelnen Ausprägungen zur *TinkerPlots-Kompetenz*, jeweils mit Definition und Ankerbeispiel, zu finden.

2.3.2 Kategoriensystem Verteilungsvergleichskompetenz

Um die *Verteilungsvergleichskompetenz* zu operationalisieren, wurden zunächst auf einer strukturierenden Stufe mögliche tragfähige Verteilungsvergleichselemente (siehe Abschnitt 2.1: *Zentrum, Streuung, Verschiebung, Form, Vergleich p-basiert, Vergleich q-basiert*) herausgearbeitet und in einem zweiten Schritt für jedes Verteilungsvergleichselement Kompetenzausprägungen definiert. Die Kategorien von Pfannkuch (2007) mit den Ausprägungen *describer*, *decoder* und *assessor* wurden zu zwei Ausprägungen zusammengefasst: *Verteilungsvergleichskompetenz* hoch und *Verteilungsvergleichskompetenz* mittel. Eine mittlere *Verteilungsvergleichskompetenz* ist gegeben, wenn Unterschiede zwischen zwei Verteilungen lediglich auf der beschreibenden Ebene herausgearbeitet werden (*describer*). Am Beispiel des Vergleichs der Zentren zweier Verteilungen bedeutet das, dass beispielsweise der Unterschied zwischen arithmetischen Mittelwerten lediglich beschrieben, aber nicht interpretiert wird („Das arithmetische Mittel der Männer ist um 833 € höher als das bei den Frauen“). Eine hohe Teilkompetenz in Pfannkuchs (2007) Sinne (*assessor*) ist gegeben, wenn die Unterschiede zwischen den arithmetischen Mittelwerten interpretiert werden („Die Männer verdienen durchschnittlich 833 € mehr als die Frauen“). Eine niedrige Teilkompetenz ist gegeben, wenn die Unterschiede inkorrekt oder fehlerhaft herausgearbeitet werden. Zusammenfassend sieht man in der Tabelle 6 im Anhang die Definitionen der einzelnen Ausprägungen zur *Verteilungsvergleichskompetenz*. Die zugehörigen Ankerbeispiele finden sich in Tabelle 7 im Anhang.

Als Fazit dieses Abschnitts werden mit Blick auf die Forschungsfrage für die explorative Fallstudie die folgenden Unterforschungsfragen abgeleitet, die im Rahmen dieser Fallstudie als gleichberechtigt betrachtet werden:

- Wie vergleichen Studierende des Lehramts Mathematik GHRGe zwei gegebene Verteilungen eines numerischen Merkmals auf der inhaltlichen Ebene der Stochastik? Welche der sechs Verteilungsvergleichselemente (*Zentrum, Streuung, Verschiebung, Form, p-basiert* und *q-basiert*) nutzen sie wie?
- Wie setzen Studierende des Lehramts Mathematik GHRGe die Software TinkerPlots in ihrem Untersuchungsprozess ein und inwiefern können Studierende des Lehramts Mathematik GHRGe mit ihrer Hilfe Verteilungsvergleiche durchführen?

3. Design und Durchführung einer explorativen Fallstudie

Ausgewählte Teilnehmer aus der Lehrveranstaltung „Statistisch denken und forschen lernen mit der Software TinkerPlots“ haben, ca. vier Wochen nachdem sie diese besucht haben, an der explorativen Fallstudie teilgenommen. Die Lehrveranstaltung wurde auf Grundlage der Forderungen für *Statistical Reasoning Learning Environments* (nach Cobb & McClain, 2004 und Garfield & Ben-Zvi, 2008) konzipiert und fundamentale inhaltliche Ideen aus der Perspektive der internationalen Stochastikdidaktik, wie u. a. das Durchlaufen des vorgenannten Datenanalyse-Zyklus PPDAC, (Wild & Pfannkuch 1999), das Arbeiten mit realen und multivariaten Daten, der Einsatz adäquater Software und Aktivitäten wie Verteilungen vergleichen wurden in die Lehrveranstaltung implementiert. Als Veranstaltungsform wurde das Seminar gewählt, weil in dieser Lehrform im Gegensatz zur Vorlesung eher Partner- und Gruppenarbeit möglich sind. Insgesamt bestand das Seminar aus 14 Seminarsitzungen, die jeweils 90 Minuten dauerten. Am Seminar haben 22 Studierende teilgenommen. Davon haben 15 Studierende Lehramt Mathematik für Grundschule und 7 Studierende Lehramt Mathematik für Haupt-, Real- und Gesamtschule studiert. Die Semesterzahl variierte zwischen dem vierten und elften Semester und alle Teilnehmer haben die Grundlagenveranstaltung „Elemente der Stochastik“ besucht. Die inhaltliche Umsetzung des Seminars wurde anhand von vier Bausteinen vollzogen:

- Baustein 1: Generierung statistischer Frage- und Problemstellungen, Vorbereitung und Durchführung einer Datenerhebung, Erstellen eines Instruments zur Datenerhebung
- Baustein 2: Einführung in die Datenanalyse mit TinkerPlots: Erste Graphiken mit TinkerPlots erstellen
- Baustein 3: Vertiefung der Datenanalyse mit TinkerPlots: Beschreibung und Interpretation von Verteilungen eines numerischen Merkmals, Vergleich von Verteilungen eines numerischen Merkmals
- Baustein 4: Hinführung zur Inferenzstatistik: Randomisierungstests mit TinkerPlots.

Wie der Auflistung der Bausteine zu entnehmen ist, haben die Teilnehmer der Lehrveranstaltung zunächst selbst eine statistische Problemstellung generiert, auf dieser Grundlage eigenständig einen Fragebogen entworfen, eine Datenerhebung durchgeführt und die Daten bereinigt (Baustein 1). Im Anschluss fand die Einführung in die Datenanalyse mit der Software TinkerPlots statt. Dort haben die Studierenden

gelernt, mit ihren Daten Verteilungen von kategorialen und numerischen Variablen mit TinkerPlots zu erstellen, zu lesen und zu interpretieren (Baustein 2) und dann weiterführend auch Verteilungen zu vergleichen sowie den Zusammenhang zwischen zwei numerischen Variablen zu untersuchen (Baustein 3). Abschließend wurden die Studierenden an die Inferenzstatistik herangeführt, indem sie basierend auf ihren Verteilungsvergleichen aus Baustein 3 Randomisierungstests mit TinkerPlots durchgeführt haben, um ihre Erkenntnisse gegebenenfalls verallgemeinern zu können (Baustein 4).

Ein besonderer Fokus dieser Lehrveranstaltung liegt auf dem Vergleich von Verteilungen eines numerischen Merkmals (Baustein 3). Für den Vergleich von Verteilungen eines numerischen Merkmals wurde in der Lehrveranstaltung eine Norm auf Grundlage der Vorarbeiten und der bestehenden Forschungsergebnisse (siehe 2.1 und 2.2) etabliert. Zum einen sollten zwischen Verteilungen eines numerischen Merkmals anhand der tragfähigen Konzepte *Zentrum, Streuung, Form, Verschiebung, p-basiert und q-basiert* so viele Unterschiede mit TinkerPlots wie möglich herausgearbeitet werden, zum anderen sollten diese herausgearbeiteten Unterschiede nicht nur beschrieben, sondern auch im Datenkontext interpretiert werden. Weitergehende Details zum Design, zur Durchführung und zur Evaluation des gesamten Seminars finden sich in Frischmeier (2017b).

3.1 Design der Fallstudie

In der Fallstudie sollte eine Verteilungsvergleichsaufgabe mit TinkerPlots unter Laborbedingungen in Zweierteams eigenständig durchgeführt werden. In der Arbeitsphase wurden die Teilnehmer aufgefordert, in Partnerarbeit gemeinsam an der Aufgabe zu arbeiten (siehe Abb. 2). Es gab keine Intervention seitens des Interviewers. Die Teilnehmer wurden ermutigt, bei der Bearbeitung der Aufgabe mit ihrem Partner zu reden und zu diskutieren und ihre Gedanken frei zu äußern.



Abb. 2: Studierende bei der Bearbeitung der Aufgabe zur Verdienststrukturerhebung

Die Softwareaktivitäten sowie die Kommunikation der Teilnehmer untereinander wurden mit der Bildschirmaufzeichnungssoftware Camtasia sowie mit einer externen Videokamera aufgezeichnet. Die Aufgabe, die in Partnerarbeit zu bearbeiten war, sollte die Auseinandersetzung mit einem komplexen Datensatz beinhalten und sollte, wie von Garfield und Ben-Zvi (2008) vorgeschlagen, motivierend, real und alltagsnah für die Studierenden sein. Das Thema Einkommensunterschiede zwischen männlichen und weiblichen Arbeitnehmern sowie der Datensatz zur Verdienststrukturerhebung 2006 des statistischen Bundesamts erfüllen diese Forderungen. Dieser Datensatz enthält in seiner ursprünglichen Form über 60.000 Fälle und mehr als 20 Variablen. In der Lehrveranstaltung wurde vorwiegend mit kleineren multivariaten Datensätzen, die jeweils unter 1000 Fälle umfassten, gearbeitet. Daher wurde in der Interviewstudie lediglich eine reduzierte Stichprobe der Verdienststrukturerhebung ($n=861$) verwendet, weil der Umfang des Datensatzes somit näher an den Datensätzen lag, die in der Lehrveranstaltung verwendet worden sind.

3.2 Die Aufgabe zur Verdienststrukturerhebung 2006

Der Datensatz zur Verdienststrukturerhebung 2006 wurde von der Homepage des Statistischen Bundesamts⁴ in die Software TinkerPlots importiert und als TinkerPlots-Datensatz zusammen mit einer Variablenliste und einem Arbeitszettel, auf dem Notizen festgehalten werden konnten, den Teilnehmern zur Verfügung gestellt. Die Aufgabenstellung zielte auf das Herausarbeiten von Verdienstunterschieden zwischen männlichen und weiblichen Arbeitnehmern ab. Die konkrete Aufgabenstellung in Form eines Vergleichungsvergleiches lautete wie folgt: „Inwiefern unterscheiden sich die Arbeitnehmerinnen und Arbeitnehmer hinsichtlich ihres Bruttomonatsgehalts? Arbeiten Sie Unterschiede in beiden Verteilungen heraus.“

3.2.1 Mögliche Bearbeitungsweisen der Aufgabe zur Verdienststrukturerhebung

Im Folgenden werden mögliche Lösungsalternativen zur Bearbeitung der Aufgabe aufgezeigt. Für den Vergleich von Verteilungen eines numerischen Merkmals wurde in der Lehrveranstaltung „Statistisch denken und forschen lernen mit der Software TinkerPlots“, wie im obigen Abschnitt berichtet wurde, eine gemeinsame Norm etabliert. Zum einen sollten zwischen zwei Verteilungen eines numerischen Merkmals anhand der tragfähigen Konzepte *Zentrum*, *Streuung*, *Form*, *Verschiebung*, *p-basiert* und *q-basiert* so viele Unterschiede wie möglich her-

ausgearbeitet werden und zum anderen diese herausgearbeiteten Unterschiede auch im Kontext interpretiert werden. Dabei konnten „Standardgraphiken“ in TinkerPlots wie gestapelte Punktdiagramme, Histogramme und Boxplots erstellt und entsprechend in dieser Form verglichen und Unterschiede zwischen den Verteilungen mithilfe der Software TinkerPlots auf vielfältige Weise herausgearbeitet werden. Es wurde erwartet, dass die Teilnehmer anhand von den im Seminar als geeignet bezeichneten TinkerPlots-Darstellungen (gestapeltes Punktdiagramm, Boxplot, Histogramm) möglichst viele Unterschiede zwischen beiden Verteilungen herausarbeiten und diese Unterschiede im Datenkontext auch interpretieren. Exemplarisch werden im Folgenden einige dieser Vergleiche herausgearbeitet. In Abb. 3, 4 und 5 sehen wir verschiedene Auswertungsgraphiken, die die Teilnehmer bei der Bearbeitung dieser Aufgabe mit TinkerPlots unter anderem erstellen konnten. Eine erste Möglichkeit ist gegeben, die Verteilungen anhand ihrer Mittelwerte zu vergleichen. Berechnet man die arithmetischen Mittelwerte der beiden Verteilungen (siehe eingezeichnete Werte in Abb. 3), so lässt sich sagen, dass das arithmetische Mittel der Verteilung des Merkmals Bruttomonatsverdienst der männlichen Arbeitnehmer um 832,80 € höher ist als das der weiblichen Arbeitnehmer. Das wäre eine Vergleichsaussage, die auf einer deskriptiven Ebene vollzogen wird: Auf dieser Ebene wird lediglich der Unterschied als Differenz der arithmetischen Mittelwerte herausgearbeitet. Eine höher einzustufende Aussage, die bereits eine Interpretation der Unterschiede der arithmetischen Mittelwerte der beiden Verteilungen im Datenkontext enthält, wäre z. B., dass die männlichen Arbeitnehmer im Durchschnitt 832,80 € mehr im Monat verdienen als die weiblichen Arbeitnehmer. Ähnliche Aussagen lassen sich auch mit den Medianen der beiden Verteilungen machen (Abb. 3). Hier könnte man auf einer beschreibenden, deskriptiven Ebene zunächst feststellen, dass der Median der Verteilung des Merkmals Bruttomonatsverdienst bei den männlichen Arbeitnehmern um 813,50 € höher ist als bei den weiblichen Arbeitnehmern. Eine höher zu bewertende Aussage würde eine Interpretation dieses Sachverhalts beinhalten, wie z. B. im Median verdienen die männlichen Arbeitnehmer 813,50 € mehr als die weiblichen Arbeitnehmer. Mithilfe von Boxplots (Abb. 3) lassen sich weitere Unterschiede zwischen den Verteilungen herausarbeiten. So lässt sich unter anderem eine Verschiebung zwischen beiden Verteilungen ausmachen. Diese lässt sich präzisieren, indem man feststellt, dass z. B. der Median der Verteilung des Merkmals Bruttomonatsverdienst bei den männlichen Arbeitnehmern genauso groß ist wie das dritte Quartil der Verteilung des Merkmals Bruttomonatsverdienst bei den weiblichen Arbeitnehmern.

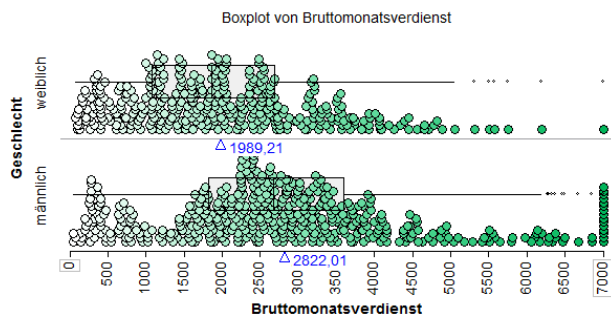


Abb. 3: Mögliche TinkerPlots Graphik „Boxplots“

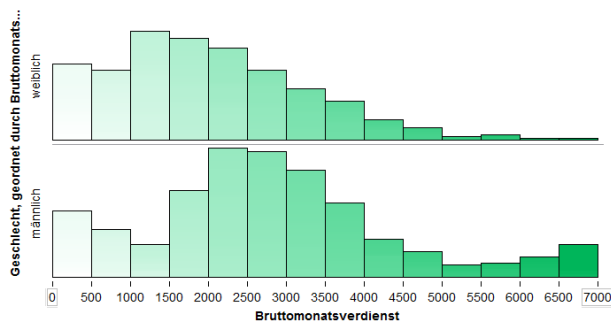


Abb. 4: Mögliche TinkerPlots Graphik „Histogramme“

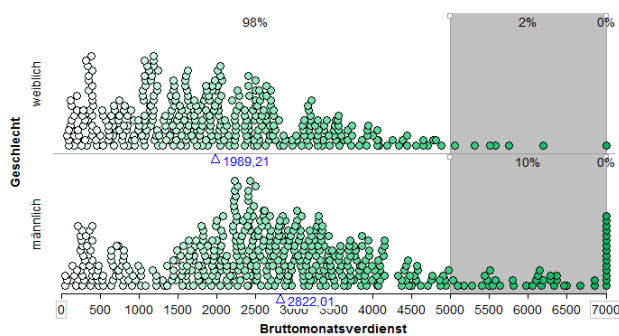


Abb. 5: Mögliche TinkerPlots Graphik „gestapelte Punktdiagramme“

Ebenso lässt sich feststellen, dass der Median der Verteilung des Merkmals Bruttomonatsverdienst bei den weiblichen Arbeitnehmern ungefähr gleich dem ersten Quartil der Verteilung des Merkmals Bruttomonatsverdienst bei den männlichen Arbeitnehmern ist. Dieses könnten Lernende so versprachlichen: Ca. 50 % der Arbeitnehmerinnen verdienen weniger als 1840 €, bei den männlichen Arbeitnehmern beträgt dieser Anteil gerade einmal ca. 25 %. Durch den Vergleich der Interquartilsabstände lassen sich dann auch Unterschiede im Hinblick auf die Streuung zwischen den beiden Verteilungen herausarbeiten: Der Interquartilsabstand beträgt bei der Verteilung des Merkmals Bruttomonatsverdienst bei den männlichen Arbeitnehmern 1767 € und bei den weiblichen Arbeitnehmern 1611,50 €. Eine erste deskriptive Beobachtung wäre, dass der Interquartilsabstand in der Gruppe der männlichen Arbeitnehmer größer ist als der Interquartilsabstand in der Gruppe der weiblichen Arbeitnehmer. Auf dieser Grundlage könnte man dann auf einer höheren Ebene festhalten, dass die Streuung in der Gruppe der Arbeitnehmer ein wenig

größer ist und das Gehaltsgefüge in dieser Stichprobe somit heterogener ist als im Vergleich zu den weiblichen Arbeitnehmern. Einen weiteren größeren Streuungsunterschied zwischen den beiden Verteilungen könnten Lernende herausarbeiten, wenn sie die Längen der linken Antennen der beiden Boxplots der beiden Verteilungen vergleichen. Anhand von Histogrammen (z. B. Abb. 4) können Unterschiede zwischen der Form zweier Verteilungen herausgearbeitet werden. Hier würden das Variieren verschiedener Klassenbreiten und der Vergleich der Verteilungen anhand der Formen zu verschiedenen Einsichten führen wie etwa, dass die Verteilung des Merkmals Bruttomonatsverdienst der weiblichen Arbeitnehmer linkssteil ist und dass sich in der Verteilung des Merkmals Bruttomonatsverdienst der Arbeitnehmer verschiedene Verteilungsabschnitte, wie z. B. $[0; 2000 \text{ €}]$, $[2000 \text{ €}; 4500 \text{ €}]$ und $[4500 \text{ €}; 7000 \text{ €}]$ identifizieren lassen. In diesen Abschnitten, wie z. B. im Intervall $[2000 \text{ €}; 4500 \text{ €}]$ lässt sich dann anhand der Histogramme wiederum eine linkssteile Verteilung erkennen. Schließlich – bezogen auf die Verteilungsvergleichselemente in Abschnitt 2.1 – können Unterschiede anhand von p- und q-basierten Vergleichen durchgeführt werden. Man könnte z. B. Vielverdiener als Personen definieren, die 5000 € oder mehr im Monat verdienen. Dann könnte man anhand der Abbildung 5 feststellen, dass bei den männlichen Arbeitnehmern der Anteil der Vielverdiener bei 10 %, bei den weiblichen Arbeitnehmern der Anteil der Vielverdiener bei nur 2 % liegt. Die einzelnen Vergleichsmöglichkeiten sowie mögliche Vorgehensweisen der Probanden beim Bearbeiten dieser Aufgabe sollten hier nur angedeutet werden, eine ausführliche Aufgabenanalyse zum Vergleich dieser Verteilungen findet sich in Biehler und Frischemeier (2015) und Frischemeier (2017b, S. 441 ff.).

3.3 Teilnehmer und Daten

Insgesamt haben acht Teilnehmer, die auch die oben beschriebene Lehrveranstaltung „Statistisch denken und forschen lernen“ besucht haben, an der Laborstudie teilgenommen. Die Teilnahme am Interview war freiwillig, die Teilnehmer wurden damit willkürlich, aber nicht streng zufällig bestimmt. Die Aufgabenbearbeitung und Interviewsituation fand unter Laborbedingungen in den Räumen der Universität Paderborn statt. Dabei wurden die Teilnehmer aufgefordert, während der Aufgabenbearbeitung miteinander zu kommunizieren. Als Material haben sie einen Arbeitszettel mit der Aufgabenstellung sowie eine TinkerPlots-Datei mit dem entsprechenden Datensatz der Verdienststrukturerhebung 2006 sowie eine Variablenliste zum Datensatz ausgehändigt bekommen.

Als Daten zur Studie wurden die Bearbeitungen der Arbeitsphase auf Video aufgenommen sowie die

D. Frischemeier

Bildschirmaktivitäten und die Kommunikation mit Camtasia aufgezeichnet. Weiterhin wurden die TinkerPlots-Dateien der Probanden sowie die Mitschriften auf dem Arbeitszettel eingesammelt.

Die Kommunikation der Teilnehmer und Aktionen mit der Software wurden transkribiert und für die weitere Analyse nach Kuckartz (2012) aufbereitet. Die Analyse der Transkripte wurde anhand der strukturierenden (qualitativen) Inhaltsanalyse nach Mayring (2010) durchgeführt.

3.4 Auswertungsmethodik

Das Vorgehen der Auswertung der Transkripte mit der qualitativen Inhaltsanalyse lief (adaptiert nach Mayring 2010) in den folgenden Schritten ab, wie sie der Abbildung 6 zu entnehmen sind.

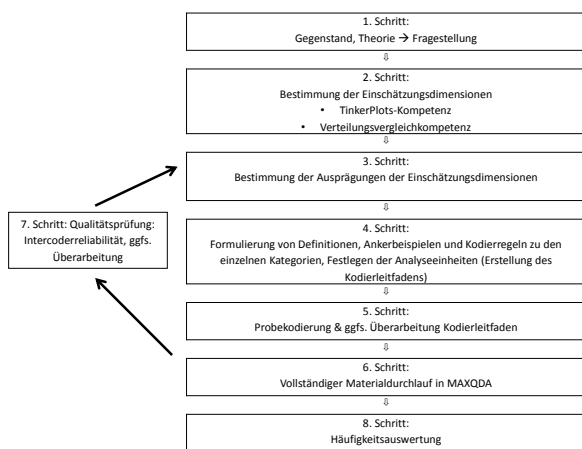


Abb. 6: Ablauf der qualitativen Inhaltsanalyse in der vorliegenden Studie (adaptiert nach Mayring 2010)

Zunächst (1. Schritt) wurde der Forschungsgegenstand (Aufarbeitung der Theorie, siehe Ausführungen in 2.1 und 2.2) festgelegt und eine Forschungsfragestellung („Wie vergleichen Studierende des Lehramts Mathematik GHRGe zwei gegebene Verteilungen eines numerischen Merkmals mit der Software TinkerPlots?“) generiert. Nach dem Aufbereiten der Daten in Form von Transkripten, die Kommunikation und Softwareaktivitäten enthalten, wurden in einem zweiten Schritt die Einschätzungsdimensionen (*TinkerPlots-Kompetenz*, *Verteilungsvergleichskompetenz*) festgelegt. Um die Softwarekompetenzen und die stochastischen Kompetenzen der Studierenden einzuschätzen, wurden im dritten Schritt die Skalierungen zu den Einschätzungsdimensionen eingefügt. Der vierte Auswertungsschritt sieht die Festlegung von Kodier- und Analyseeinheiten sowie die Erstellung der Kategorien samt Definitionen und Ankerbeispielen vor. Als Analyseeinheit wurden bei der Dimension *TinkerPlots-Kompetenz* die Camtasiaaufnahmen und die Videoaufnahmen sowie die Transkripte festgelegt. Bei der Dimension *Verteilungsvergleichs-*

kompetenz umfasste die Analyseeinheit die Camtasia-Aufnahmen und Videoaufnahmen, die Transkripte sowie die Mitschriften auf den Arbeitszetteln. Beim Kategoriensystem *TinkerPlots-Kompetenz* wurde als minimale Kodiereinheit ein Wort und als maximale Kodiereinheit einen Sinnabschnitt ausgewählt. Beim Kategoriensystem *Verteilungsvergleichskompetenz* ist die minimale Kodiereinheit ein Wort und die maximale Kodiereinheit eine Phrase/Aussageeinheit. Die Bildung der Kategorien wurde bereits in 2.3 beschrieben, die entsprechenden Kodiermanuale sind im Anhang (Tabellen 5-7) zu finden. Im Anschluss wurde der Kodierleitfaden erstellt und dann eine Probekodierung (5. Schritt) vorgenommen. Nach einer Überarbeitung des Kodierleitfadens wurde dann nach einem Materialdurchlauf (6. Schritt) die Prüfung der Interkoderreliabilität (7. Schritt) vorgenommen. Die Interkoderreliabilität ($\kappa = 0,8558$ für die Dimension *TinkerPlots-Kompetenz* und $\kappa = 1,000$ für die Dimension *Verteilungsvergleichskompetenz*) wurde durch die Überprüfung mit einem unabhängigen Zweitkodierer sichergestellt. Anschließend fand im letzten Schritt (8. Schritt) ein vollständiger Materialdurchlauf in der qualitativen Datenanalysesoftware *MAXQDA* statt. Mit der Software *MAXQDA* wurden dann auch die Häufigkeitsauswertungen vorgenommen, welche in den folgenden Abschnitten 4.1 und 4.2 präsentiert werden.

4. Ergebnisse

4.1 Häufigkeitsanalysen zur *TinkerPlots-Kompetenz*

Wie man der Häufigkeitsauswertung in Tabelle 1 ansehen kann, zeigen die Paare generell eine durchweg kompetente Nutzung der Software TinkerPlots beim Verteilungsvergleich, denn 75,9 % der Kodierungen zur Dimension *TinkerPlots-Kompetenz* sind einer hohen Ausprägung zugeordnet. Fasst man die hohen und die mittleren Ausprägungen zusammen, so lässt sich sogar sagen, dass ca. 86 % der Kodierungen zur *TinkerPlots-Kompetenz* eine mindestens mittlere Kompetenz aufweisen. Zwischen den einzelnen Paare lassen sich einige Unterschiede erkennen: Während Hilde und Iris (Namen geändert) eine sehr kompetente Softwarenutzung zeigen (ca. 98 % der Kodierungen zur Dimension *TinkerPlots-Kompetenz* sind einer mindestens mittleren *TinkerPlots-Kompetenz* zugeordnet), zeigen die Paare Conrad und Maria (Namen geändert) sowie Laura und Ricarda eine kompetente Softwarenutzung, die dennoch teilweise einige Unzulänglichkeiten aufweist: bei Conrad und Maria sind ca. 21 %, bei Laura und Ricarda ca. 11 % der Kodierungen einer niedrigen *TinkerPlots-Kompetenz* zugeordnet.

Paar	TP Komp. hoch	TP Komp. mittel	TP Komp. niedrig	Gesamt
Conrad & Maria	75,9 %	3,4 %	20,7 %	100,0 %
Hilde & Iris	80,5 %	17,1 %	2,4 %	100,0 %
Laura & Ricarda	81,5 %	7,4 %	11,1 %	100,0 %
Sandra & Luzie	45,5 %	9,0 %	45,5 %	100,0 %
Gesamt	75,9 %	10,2 %	13,9 %	100,0 %

Tab. 1: Verteilung der Kodierungen zur Einschätzungsdimension "TinkerPlots-Kompetenz" getrennt nach den Teilnehmerpaaren

Das Paar Sandra und Luzie (Namen geändert) zeigt im Vergleich zu den anderen drei Paaren die schlechteste Softwarekompetenz (nur ca. 45,5 % ihrer Kodierungen sind einer hohen *TinkerPlots-Kompetenz* und 45,5 % der Kodierungen sind einer niedrigen Kompetenz zugeordnet).

Zusammenfassend lässt sich auf Basis der Häufigkeitsauswertungen sagen, dass die Probanden (mit der Ausnahme von Sandra und Luzie) durchweg kompetent die Software TinkerPlots zum Verteilungsvergleich nutzen können.

4.2 Häufigkeitsanalysen zur Verteilungsvergleichskompetenz

In Tabelle 2 ist die Häufigkeitsauswertung zur Einschätzungsdimension *Verteilungsvergleichskompetenz* dargestellt. Von den 28 Kodierungen in dieser Dimension sind ungefähr 29 % einer hohen und ca. 71 % einer mittleren *Verteilungsvergleichskompetenz* zugeordnet. Es lässt sich erkennen, dass es keine Kodierungen im Bereich einer niedrigen *Verteilungsvergleichskompetenz* gibt. Somit ist schon einmal positiv festzuhalten, dass von den Teilnehmern keine inkorrekten Unterschiede zwischen den Verteilungen herausgearbeitet wurden. Das gibt Anlass zu der Annahme, dass die Probanden nach dem Seminar durchaus in der Lage sind, Verteilungen mit tragfähigen Vergleichsansätzen unter Nutzung der Software TinkerPlots kompetent zu vergleichen. Viele Vergleiche sind lediglich mit einer mittleren Kompetenz bewertet, weil die Unterschiede beschrieben aber nicht interpretiert wurden. Das deckt sich unter anderem mit den Beobachtungen von Biehler (1997), Pfannkuch et al. (2004) und Pfannkuch (2007).

So hätten viele Vergleiche, die in die Kategorie „*Verteilungsvergleichskompetenz* mittel“ eingeordnet wurden, durch eine zusätzliche Interpretation im Datenkontext verbessert werden können.

Paar	Vvgl. Komp. hoch	Vvgl. Komp. mittel	Vvgl. Komp. niedrig	Gesamt
Conrad & Maria	0 (0 %)	4 (100 %)	0 (0 %)	4 (100 %)
Hilde & Iris	2 (20 %)	8 (80 %)	0 (0 %)	10 (100 %)
Laura & Ricarda	6 (60 %)	4 (40 %)	0 (0 %)	10 (100 %)
Sandra & Luzie	0 (0 %)	4 (100 %)	0 (0 %)	4 (100 %)
Gesamt	8 (28,6 %)	20 (71,4 %)	0 (0 %)	28 (100 %)

Tab. 2: Verteilung der Kodierungen zur Einschätzungsdimension "Verteilungsvergleichskompetenz" (hier kurz: vgl. Komp.) getrennt nach den Teilnehmerpaaren

Zwischen den Paaren lassen sich weitere Unterschiede feststellen. Zum einen haben die Paare Hilde & Iris (10) und Laura & Ricarda (10) vergleichsweise durchaus mehr Unterschiede zwischen den Verteilungen herausgearbeitet als Conrad und Maria (4) oder Sandra und Luzie (4). Ebenso finden sich Unterschiede im Hinblick auf die Qualität der Vergleichsaussagen. Während bei Laura und Ricarda mehr als die Hälfte aller Vergleiche (60 %) auf eine hohe *Verteilungsvergleichskompetenz* schließen lassen, sind es bei Hilde und Iris immerhin noch 20 % der Vergleiche, auf die selbiges zutrifft. Bei Sandra und Luzie sowie bei Conrad und Maria ist kein Vergleich einer hohen Kompetenz zugeordnet. Bei diesen Paaren besteht also durchaus Verbesserungsbedarf was die Interpretation der Unterschiede angeht.

Im Folgenden werden jeweils zwei Vergleiche aus den Bearbeitungen der Teilnehmer vorgestellt, die in die Kategorie „*Verteilungsvergleichskompetenz* hoch“ und zwei Vergleiche, die in die Kategorie „*Verteilungsvergleichskompetenz* mittel“ eingeordnet werden können.

Conrad und Maria erstellen Boxplots in TinkerPlots und berechnen die arithmetischen Mittelwerte der beiden Verteilungen des Merkmals Bruttomonatsverdienst (Abb. 7). Daraufhin vergleichen sie die Lage der beiden Verteilungen anhand der Boxplots. Indem Conrad nur formuliert „[...] dass es halt weiter nach rechts verschoben ist.“ beschreibt er lediglich die Verschiebung, geht aber auf eine mögliche Interpretation des Unterschiedes nicht ein, weshalb dieser Abschnitt als mittlere *Verteilungsvergleichskompetenz* (Verschiebung) eingeordnet wurde.

D. Frischemeier

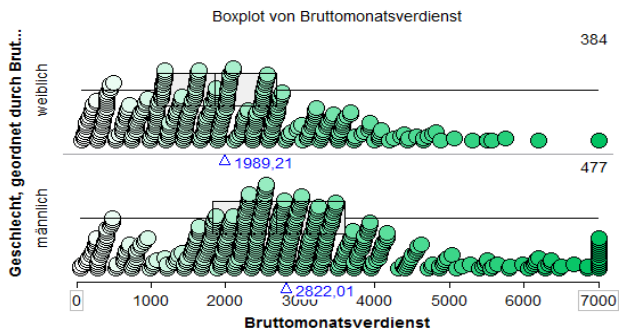


Abb. 7: TinkerPlots-Graphik von Conrad und Maria während ihrer Bearbeitung

Ein weiteres Beispiel für eine mittlere *Verteilungsvergleichskompetenz* ist unter anderem bei der Bearbeitung von Sandra und Luzie zu finden: Sandra und Luzie bestimmen mithilfe der Einteiler-Funktion in TinkerPlots (Abb. 8) die relativen Häufigkeiten der Arbeitnehmerinnen und Arbeitnehmer in den Intervallen [0 €; 2000 €[, [2000 €; 5000 €[und [5000 €; 7000 €[. Mit diesen Informationen führen sie p-basierte Vergleiche durch.

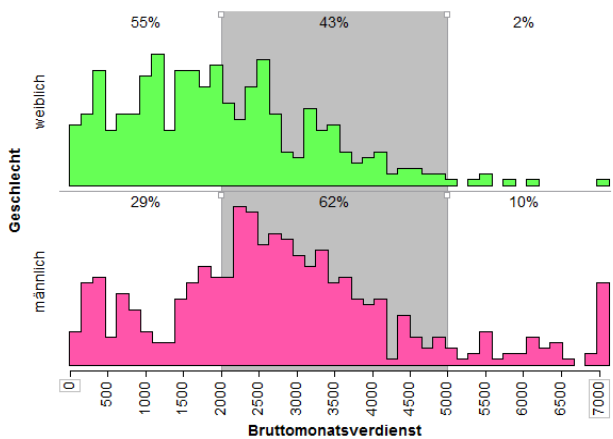


Abb. 8: TinkerPlots-Graphik von Sandra und Luzie während ihrer Bearbeitung

	Männer	Frauen
Vielverdiener: 5000 bis 7000€	10%	2%
Mittelverdiener: 2000 bis 5000€	62%	43%
Wenigverdiener: weniger als 2000	29%	55%

Abb. 9: handschriftliche Notiz von Sandra und Luzie während ihrer Bearbeitung

Sandra und Luzie lesen die entsprechenden relativen Häufigkeiten ab und dokumentieren diese handschriftlich in einer Tabelle (Abb. 9). Da die p-basierten Vergleiche durch die Dokumentation in der handschriftlichen Tabelle in Abbildung 8 lediglich angedeutet aber die Unterschiede nicht interpretiert und direkt verglichen werden, wurde diese Fundstelle als mittlere *Verteilungsvergleichskompetenz* p-basiert eingeschätzt.

Zwei Beispiele, in denen eine hohe *Verteilungsvergleichskompetenz* ersichtlich wird, seien im Folgenden illustriert. Laura und Ricarda bestimmen die arithmetischen Mittelwerte der Verteilungen des Merkmals Bruttomonatsverdienst bei den männlichen und bei den weiblichen Arbeitnehmern (Abb. 10).

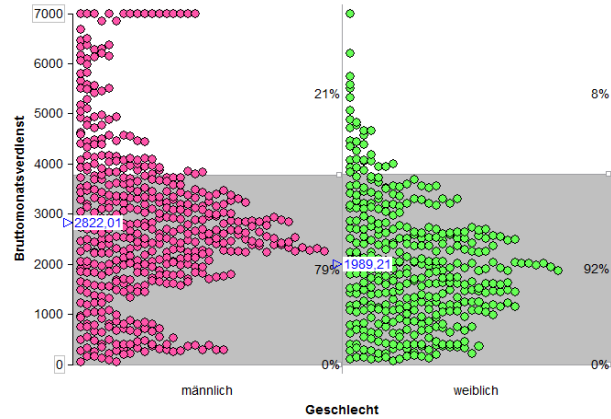


Abb. 10: TinkerPlots-Graphik von Laura und Ricarda während ihrer Bearbeitung

Aus den handschriftlichen Notizen (Abb. 11) wird ersichtlich, dass sie den multiplikativen Unterschied (29,5 %) zwischen den arithmetischen Mittelwerten berechnen und diesen im Kontext interpretieren („Die Männer verdienen ... mehr als die Frauen“).

Vergleich: Die Männer verdienen im Jahr 2006 durchschnittlich 29,5% mehr als die Frauen.

Abb. 11: Handschriftliche Notiz von Laura und Ricarda während ihrer Bearbeitung

Ein anderes Beispiel, in dem eine hohe *Verteilungsvergleichskompetenz* vorliegt (am Beispiel der Unterschiede bezüglich der Streuung) findet sich z. B. beim Paar Hilde und Iris. Hilde und Iris haben in TinkerPlots Boxplots über die beiden gestapelten Punktdiagramme (Abb. 12) gelegt und mit Hilfslinien die Werte jeweils für den Median, für das erste Quartil und für das dritte Quartil bestimmt.

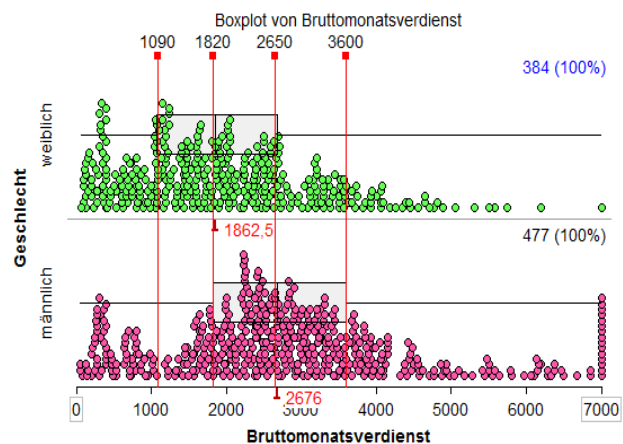


Abb. 12: TinkerPlots-Graphik von Hilde und Iris während ihrer Bearbeitung

Ausgehend von Abbildung 12 entstand nach der Erstellung der TinkerPlots-Graphik der folgende Dialog zwischen Hilde und Iris.

Iris: Ja, da sehen wir ja auch schon, dass der Interquartilsabstand ... bei den 1600 liegt [deutet auf die Verteilung des Merkmals Bruttomonatsverdienst bei den Arbeitnehmerinnen] und hier dann bei 1800 [deutet auf die Verteilung des Merkmals Bruttomonatsverdienst bei den Arbeitnehmern]

Hilde: Das heißt also, dass die Streuung von dem-, ähm (nachdenkend) ... mittleren 50 %

Iris: streut mehr

Hilde: streut mehr bei den Männern

Im Dialog wird ersichtlich, dass Hilde und Iris zunächst bei beiden Verteilungen die Interquartilsabstände berechnen. Im Anschluss vergleichen sie diese dann und interpretieren die Unterschiede der Interquartilsabstände korrekt („streut mehr bei den Männern“). Auch dieses Beispiel wurde eingeordnet in die Kategorie „hohe *Verteilungsvergleichskompetenz*“ (hinsichtlich des Vergleichs der Streuung).

Der Tabelle 3 kann entnommen werden, welche tragfähigen Verteilungsvergleichselemente von den vier Paaren zum Vergleich genutzt werden. Zunächst ist zu sagen, dass die vier Teilnehmerpaare eine breite Palette an Verteilungsvergleichselementen nutzen. Dies ist sehr positiv zu sehen, wenn man bedenkt, dass einige empirische Studien (z. B. Frischemeier & Biehler, 2011) eher ein „Zurückfallen auf Mittelwerte“ und den ausschließlichen Gebrauch des arithmetischen Mittels beim Vergleich von Verteilungen beobachten. Hier werden auch die Auswirkungen der Lehrveranstaltung „Statistisch denken und forschen lernen mit der Software TinkerPlots“ sichtbar, die ja gerade den Teilnehmern eine breite Palette an Verteilungsvergleichselementen zur Verfügung stellen sollte.

Es fällt auf, dass die Verteilungen bevorzugt anhand der Elemente „p-basiert“ (8mal), „Verschiebung“ (6mal) und „Streuung“ (5mal) verglichen wurden.

Vergl.	C & M	H & I	L & R	S & L	Ges.
Zentrum	0	2	2	0	4
Streuung	2	2	1	0	5
Form	0	0	3	0	3
Versch.	1	4	1	0	6
p-basiert	0	2	2	4	8
q-basiert	1	0	1	0	2
Gesamt	4	10	10	4	28

Tab. 3: Verteilung der Kodierungen der Verteilungsvergleichselemente getrennt nach den Teilnehmerpaaren (Vergl. = Vergleichselemente; Versch. = Verschiebung; C & M = Conrad & Maria; H & I = Hilde & Irene; L & R = Laura & Ricarda; S & L = Sandra & Luzie)

Die Vergleichselemente „Form“ (3mal) und „q-basiert“ (2mal) werden vergleichsweise seltener genutzt.

Es stellt sich die Frage, welche Verteilungsvergleichselemente mit welcher Qualität genutzt werden. Beim Blick auf Tabelle 4 fällt auf, dass bezüglich des Verteilungsvergleichselements „Verschiebung“ und bezüglich „p-basierter Vergleiche“ kein Vergleich in die Kategorie „hohe Kompetenz“ eingeordnet ist.

Vergl.	Vvgl. Komp. hoch	Vvgl. Komp. mittel	Vvgl. Komp. niedrig	Gesamt
Zentrum	2	2	0	4
Streuung	3	2	0	5
Form	2	1	0	3
Versch.	0	6	0	6
p-basiert	0	8	0	8
q-basiert	1	1	0	2
Gesamt	8	20	0	28

Tab. 4: Verteilung der Kodierungen zur Einschätzungsdimension „Verteilungsvergleichskompetenz“ in Bezug auf die Verteilungsvergleichselemente

Dies könnte so gedeutet werden, dass die Interpretation bezüglich von Unterschieden der Verschiebung sowie p-basierte Unterschiede zweier Verteilungen eines numerischen Merkmals den Teilnehmern Schwierigkeiten bereitet. Einzig bei den Vergleichen bezüglich der Streuung sowie bezüglich der Form dominieren die Kodierungen zu einer hohen Kompetenz, bei den Vergleichen bezüglich des Zentrums und bei q-basierten Vergleichen ist die Situation ausgeglichen.

5. Zusammenfassung & Ausblick

5.1 Zusammenfassung und Diskussion der Ergebnisse

Die explorative Fallstudie zeigt, dass die Teilnehmer, die an der Lehrveranstaltung „Statistisch denken und forschen lernen mit der Software TinkerPlots“ teilgenommen haben, eigenständig in der Lage sind, in einem großen und realen Datensatz, wie z. B. dem Datensatz zur Verdienststrukturerhebung 2006, nach vorgegebenen Fragestellungen mit TinkerPlots Unterschiede zwischen zwei Verteilungen eines numerischen Merkmals herauszuarbeiten. Dabei fallen die Teilnehmer nicht auf Mittelwerte (wie z. B. bei Frischemeier & Biehler, 2011) zurück, sondern nutzen eine breite Palette der sechs in Abschnitt 2.1 als tragfähig identifizierten Verteilungsvergleichselemente, um Unterschiede zwischen den Verteilungen herauszuarbeiten. In ihrem Vorgehen zeigen die Teilnehmer durchweg eine mindestens mittlere *Verteilungsvergleichskompetenz*, allerdings werden die herausgearbeiteten Unterschiede oftmals nicht inter-

D. Frischemeier

pretiert. Dieses findet sich auch in anderen empirischen Untersuchungen (u. a. siehe auch Biehler, 1997; Pfannkuch et al., 2004; Pfannkuch, 2007), in denen berichtet wird, dass Lernende beim Verteilungsvergleich oft nur auf der „beschreibenden“ Ebene bleiben, aber die Unterschiede zwischen den Verteilungen nicht interpretieren.

Außerdem lässt sich als ein wesentliches Resultat der Fallstudie festhalten, dass die Software TinkerPlots durchgängig (mit teilweisen Einschränkungen beim Paar Sandra & Luzie) kompetent von den Teilnehmern genutzt wurde, um ihre Intentionen beim Vergleich von Verteilungen zu verfolgen und die von ihnen anvisierten Unterschiede herauszuarbeiten. Somit kann TinkerPlots eine geeignete Software im Rahmen dieser Phase der Lehrerbildung in Stochastik darstellen: die Studierenden können die Software kompetent für ihre Datenanalyse- und Verteilungsvergleichsvorhaben nutzen und lernen darüber hinaus auch eine Datenanalysesoftware kennen, die sie später im Unterricht entweder in Form eines Schülerarbeitsmediums oder als Demonstrationsmedium nutzen können. Dies ist eine wichtige Erkenntnis, denn ein tragfähiges *statistical content knowledge* sowie *technological knowledge* ist unabdingbar dafür, dass die Lehramtsstudierenden die Software und die Verteilungsvergleichsaktivitäten auch später in ihrem Unterricht einsetzen.

Weiterhin lassen sich aus den Ergebnissen der Fallstudie auch Vorschläge für das Re-Design der Lehrveranstaltung und für das Design von Verteilungsvergleichsaufgaben ableiten. So müssen einzelne Normen, die in Lehrveranstaltungen gesetzt werden, über das „Beschreiben“ hinausgehen und das Interpretieren der Unterschiede (gerade bei p-basierten Vergleichen sowie bei Vergleichen bezüglich der Verschiebung, also bei Verteilungsvergleichselementen, bei denen die Interpretation den Teilnehmern sowieso schon schwerfällt) zwischen den Verteilungen noch stärker betont werden. Vor allem müssen gute und weniger gute Beispiele für Interpretationen vorgestellt und im Unterrichtsgespräch bzw. im Plenum gezielter diskutiert werden, um diesbezüglich entsprechende Normen zu setzen.

Mit Blick auf die Lehre (in Schule und Hochschule) kann das Kategoriensystem zur *Verteilungsvergleichskompetenz* eingesetzt werden, um eine Norm zu etablieren und Lernenden verschiedene Vergleichsmöglichkeiten sowie verschiedene Stufen (beschreibend, interpretierend) von Vergleichsmöglichkeiten aufzuzeigen. Ein Plan – in Form eines Datenanalyseplans (ähnlich zu der Idee des Simulationsplans beim Simulieren von Zufallsexperimenten, siehe Biehler & Maxara, 2007) – könnte im Weiteren

ein strukturiertes Vorgehen beim Verteilungsvergleich unterstützen und darüber hinaus eine Möglichkeit zur Dokumentation der Ergebnisse (und der Unterschiede zwischen den Verteilungen) beitragen. Ebenso könnte ein solcher Plan helfen, Lernende auf mögliche Vergleichs- und mögliche Interpretationskomponenten aufmerksam zu machen.

Das Schema zur *TinkerPlots-Kompetenz* kann genutzt werden, um die Softwarekompetenz von Lernenden einzuschätzen. Das kann nicht nur bei der Verwendung der speziellen Software TinkerPlots, sondern auch bei der Verwendung von anderen digitalen Werkzeugen bei der Bearbeitung von Aufgaben aus dem Bereich der Datenanalyse geschehen.

Im Zeitalter von Big Data und frei verfügbaren öffentlichen Datensätzen im Internet, bietet die hier verwendete Aufgabe sowie der damit verbundene Datensatz des statistischen Bundesamts zur Verdienststrukturerhebung 2006 eine gute Möglichkeit, reale, bedeutungsvolle und multivariate Daten im Stochastikunterricht mit digitalen Werkzeugen wie TinkerPlots explorieren zu können (für das Beispiel der Verdienststrukturerhebung siehe auch Biehler & Frischemeier, 2015; für die Exploration gesellschaftlich relevanter Daten in anderen Themenfeldern und mit anderen Tools siehe z. B. Krüger, 2012).

Als Einschränkung bei dieser Studie ist zu betonen, dass es sich hier um eine explorative Fallstudie handelt, die nur vier Paare (bzw. acht Teilnehmer) umfasst. Die Teilnehmer hatten durch den Besuch der entsprechenden Lehrveranstaltung eine gewisse Vorerfahrung und wurden nicht streng zufällig ausgewählt, somit sind allgemeinere Schlussfolgerungen über die Studie hinaus nicht möglich.

5.2 Ausblick

Die Ergebnisse und Erkenntnisse dieser Fallstudie sind zum Anlass genommen worden, Veranstaltungen und Lernumgebungen zur Förderung der „Datenanalysekompetenz“ und das Vergleichen von Verteilungen mit TinkerPlots stetig weiter zu entwickeln. So ist eine Lehrveranstaltung zur Förderung der stochastischen Kompetenzen von Primarstufenlehramtsstudierenden an der Universität Paderborn entwickelt worden, die auch die Software TinkerPlots einsetzt und in der Verteilungsvergleiche eine besondere Rolle spielen (Podworny et al., 2017). Darüber hinaus ist außerdem ein Seminar „Statistische Allgemeinbildung im Mathematikunterricht (Sekundarstufe I)“ entwickelt worden, in welchem der hier verwendete Datensatz zur Verdienststrukturerhebung im Rahmen einer vierwöchigen Projektarbeit zum GenderPayGap in Deutschland von den Teilnehmern mit digitalen Werkzeugen exploriert wird (Frischmeier, Podworny & Biehler, 2018).

Abgrenzend von der universitären Bildung wurden in weiteren Unterrichtsprojekten Verteilungsvergleiche elementarisiert und durch Nutzung von Präkonzepten für Zentrum und Streuung wie modale Haufen (siehe z. B. Konold et al., 2002) oder Hutplots (siehe z. B. Watson, Fitzallen, Wilson & Creed, 2008) in den Projektunterricht der Primarstufe eingebunden (Breker 2016; Schäfers 2017), um somit bereits in der Grundschule eine tragfähige Datenkompetenz aufzubauen. Empirische Begleitstudien (z. B. Frischemeier, 2017a) zeigen, dass Viertklässler nach einer Unterrichtsreihe Verteilungen mit TinkerPlots unter Nutzung von Präkonzepten wie modale Haufen oder Hutplots tragfähig vergleichen können.

Weitere Forschungsaktivitäten rund um den Einsatz von TinkerPlots sowie der Förderung einer Datenkompetenz und der Entwicklung einer Verteilungsvergleichskompetenz sind momentan in Durchführung.

Anmerkungen

¹ Wird in diesem Artikel von Schülern, Studenten, Lehrern, etc. gesprochen, sind diese und entsprechende Bezeichnungen als „geschlechtsneutral“ aufzufassen. In dem vorliegenden Artikel wird auf die feminine Form ausschließlich aufgrund der besseren Lesbarkeit verzichtet.

² Für eine spezifische Unterscheidung zwischen „Knowledge of Statistics Content“, „Knowledge of Technology for doing statistics“ und „Knowledge of Pedagogy“ siehe Groth (2007).

³ Die deutsche Version der Software wurde von der Arbeitsgruppe Biehler erstellt und ist unter <https://www.tinkerplots.com/get> (aufgerufen am 15.12.2017) verfügbar.

⁴ Link Campus-file: <http://www.forschungsdatenzentrum.de/bestand/gls/cf/2006/index.asp> (aufgerufen am 19.1.2018)

Danksagung

Ein herzliches Dankeschön an Markus Vogel als betreuender Herausgeber sowie an den anonymen Reviewer für die hilfreichen Anmerkungen und Kommentare bei der Überarbeitung dieses Artikels.

Literatur

- Arbeitskreis Stochastik der Gesellschaft für Didaktik der Mathematik. (2012). Empfehlungen für die Stochastikausbildung von Lehrkräften an Grundschulen. Abgerufen von http://www.mathematik.uni-dortmund.de/ak-stoch/Empfehlungen_Stochastik_Grundschule.pdf
- Biehler, R. (1997). Students' Difficulties in Practicing Computer-Supported Data Analysis: Some Hypothetical generalizations from Results of Two Exploratory Studies. In J. Garfield & G. Burrill (Hrsg.), *Role of Technology in Teaching and Learning Statistics* (S. 169-190). Voorburg: International Statistical Institute.
- Biehler, R. (2001). Statistische Kompetenz von Schülerinnen und Schülern – Konzepte und Ergebnisse empirischer Studien am Beispiel des Vergleichens empirischer Verteilungen. In M. Borovcnik, J. Engel & D. Wickmann (Hrsg.), *Anregungen zum Stochastikunterricht* (S. 97-114). Hildesheim: Franz Becker.
- Biehler, R. (2007a). Denken in Verteilungen - Vergleichen von Verteilungen. *Der Mathematikunterricht*, 53(3), 3-11.
- Biehler, R. (2007b). Students' strategies of comparing distributions in an exploratory data analysis context. https://www.stat.auckland.ac.nz/~iase/publications/isi56/IPM37_Biehler.pdf
- Biehler, R., Ben-Zvi, D., Bakker, A. & Makar, K. (2013). Technology for Enhancing Statistical Reasoning at the School Level. In M. A. Clements, A. J. Bishop, C. Keitel-Kreidt, J. Kilpatrick & F. K.-S. Leung (Hrsg.), *Third International Handbook of Mathematics Education* (pp. 643-689). New York: Springer Science + Business Media.
- 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., Reading, C. & Shaughnessy, M. (2018). Reasoning about data In D. Ben-Zvi, K. Makar & J. Garfield (Hrsg.), *International Handbook of Research in Statistics Education* (pp. 139-192): Springer International.
- 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), 45-61.
- Biggs, J. B. & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (2006). *Bildungsstandards Mathematik: konkret*. Berlin: Cornelsen.
- Breker, R. (2016). *Design, Durchführung und Evaluation einer Unterrichtseinheit zur Entwicklung der Kompetenz „Verteilungen zu vergleichen“ in einer 4. Klasse unter Verwendung der Software TinkerPlots und neuer Medien*. (Bachelor of Education), Universität Paderborn.
- Burrill, G. & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill & C. Reading (Hrsg.), *Teaching statistics in school mathematics-Challenges*

D. Frischemeier

- for teaching and teacher education (pp. 57-69). Netherlands: Springer.
- Camtasia Studio Version 6.0.3. (2011). Okemos: Techsmith.
- Cobb, P. & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. Garfield (Hrsg.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (S. 375-396). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Fitzallen, N. & Watson, J. (2014). *Extending the curriculum with Tinkerplots: Opportunities for early development of informal inference*. Paper presented at the Ninth International Conference on Teaching Statistics, Flagstaff (AZ), USA.
- Frischemeier, D. (2017a). *Primary school students' statistical reasoning when comparing groups*. Paper presented at the IASE Satellite Conference 2017, Rabat, Morocco.
- Frischemeier, D. (2017b). *Statistisch denken und forschen lernen mit der Software TinkerPlots*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Frischemeier, D. & Biehler, R. (2011). Spielerisches Erlernen von Datenanalyse mit der Software TinkerPlots - Ergebnisse einer Pilotstudie. In R. Haug & L. Holzäpfel (Hrsg.), *Beiträge zum Mathematikunterricht 2011* (S. 275-278). Münster: WTM.
- Frischemeier, D., Podworny, S. & Biehler, R. (2018). *Activities for promoting civic statistical knowledge of preservice teachers*. Paper presented at the Challenges and Innovations in Statistics Education, Szeged.
- Garfield, J. & Ben-Zvi, D. (2008). *Developing students' statistical reasoning. Connecting Research and Teaching Practice*. The Netherlands: Springer.
- Gil, E. & Ben-Zvi, D. (2014). *Long term impact of the connections program on students' informal inferential reasoning*. Paper presented at the Ninth International Conference on Teaching Statistics, Flagstaff (AZ), USA.
- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38(5), 427-437.
- Harradine, A. & Konold, C. (2006). *How representational medium affects the data displays students make*. Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Brazil.
- Hasemann, K. & Mirwald, E. (2012). Daten, Häufigkeit und Wahrscheinlichkeit. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule: Mathematik konkret* (S. 141-161). Berlin: Cornelsen Scriptor.
- Konold, C. (2006). Designing a Data Analysis Tool for Learners. In M. Lovett & P. Shah (Hrsg.), *Thinking with data: The 33rd Annual Carnegie Symposium on Cognition*. Hillside, NJ: Lawrence Erlbaum Associates.
- Konold, C. (2007). Designing a data tool for learners. In M. Lovett & P. Shah (Hrsg.), *Thinking with data: The 33rd Annual Carnegie Symposium on Cognition* (S. 267-292). Hillside, NJ: Lawrence Erlbaum Associates.
- Konold, C. & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin & D. Schifter (Hrsg.), *A Research Companion to Principles and Standards for School Mathematics* (S. 193-215). Reston, VA: National Council of Teachers of Mathematics.
- Konold, C. & Miller, C. (2011). TinkerPlots 2.0. Emeryville, CA: Key Curriculum Press.
- Konold, C., Pollatsek, A., Well, A. & Gagnon, A. (1997). *Students Analyzing Data: Research of Critical Barriers*. In J. Garfield & G. Burrill (Hrsg.), *Research on the Role of Technology in Teaching and Learning Statistics: Proceedings of the 1996 IASE Round Table Conference* (S. 151-167). Voorburg, Netherlands: International Statistical Institute.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R. & Mayr, S. (2002). *Students' use of modal clumps to summarize data*. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
- Krüger, K. (2012). Haushaltsnettoeinkommen – ein Beispiel zur Nutzung der GENESIS-Online Datenbank im Unterricht. *Stochastik in der Schule*, 32(3), 8-14.
- Kuckartz, U. (2012). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim, Basel: Beltz Juventa.
- Makar, K. & Confrey, J. (2002). *Comparing Two Distributions: Investigating Secondary Teachers' Statistical Thinking*. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
- Makar, K. & Confrey, J. (2014). Wondering, wandering or unwavering? Learners' statistical investigations with Fathom. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen – Using Tools for Learning Mathematics and Statistics* (S. 351-362). Wiesbaden: Springer.
- Maxara, C. (2009). *Stochastische Simulation von Zufallsexperimenten mit Fathom - Eine theoretische Werkzeuganalyse und explorative Fallstudie*. Hildesheim: Franz Becker.
- Maxara, C. (2014). Konzeptualisierung unterschiedlicher Kompetenzen und ihrer Wechselwirkungen, wie sie bei der Bearbeitung von stochastischen Simulationsaufgaben mit dem Computer auftreten. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen* (S. 321-336). Wiesbaden: Springer Spektrum.
- MAXQDA 11 Software für qualitative Datenanalyse. (2013). Berlin: Sozialforschung GmbH.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Wiesbaden: Beltz.
- Mishra, P. & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *The Teachers College Record*, 108(6), 1017-1054.
- Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education*, 2(3), 149-167.
- Pfannkuch, M. & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill & C. Reading (Hrsg.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 323-333). Dordrecht/Heidelberg/London/New York: Springer.
- Pfannkuch, M., Budgett, S., Parsonage, R. & Horing, J. (2004). *Comparison of data plots: Building a pedagogical framework*. Paper presented at the Tenth International Congress on Mathematics Education (ICME-10), Copenhagen, Denmark.
- Podworny, S., Frischemeier, D. & Biehler, R. (2017). Design, Realization and Evaluation of a statistics course for preservice teachers for primary school in Germany. In A. Molnar (Hrsg.), *IASE Satellite Conference 2017: Teaching Statistics in a Data Rich World*. Rabat, Morocco: IASE.

- Rossman, A. J., Chance, B. L. & Lock, R. H. (2001). *Workshop Statistics: Discovery with data and Fathom*. Emeryville, CA: Key College Publishing.
- Rubin, A., Hammerman, J. K. L. & Konold, C. (2006). *Exploring informal inference with interactive visualization software*. Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Brazil.
- Schäfers, C. (2017). *Durchführung und qualitative Evaluation einer redesignten Unterrichtsreihe zur Entwicklung der Kompetenz "Verteilungen zu vergleichen" in einer Jahrgangsstufe 4 unter Verwendung der Software TinkerPlots*. (Bachelor of Education), University of Paderborn.
- Watson, J., Fitzallen, N., Wilson, K. & Creed, J. (2008). The representational value of HATS. *Mathematics Teaching in Middle School*, 14(1), 4-10.
- Watson, J. M. & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145-168.
- Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-265.

Anschrift des Verfassers

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

Anhang

Im Anhang finden sich die Kodiermanuale zur Fallstudie.

Tabelle 5: Ausschnitt aus dem Kodiermanual zur Einschätzungsdimension „*TinkerPlots-Kompetenz*“

Ausprägung	Definition	Ankerbeispiel
<i>TinkerPlots-Kompetenz</i> hoch	Wenn Lernende einen konkreten Plan formulieren und diesen umsetzen können, beschreiben wir die <i>TinkerPlots-Kompetenz</i> mit hoch	Conrad & Maria: “Lass uns einen Boxplot erstellen”. <i>Conrad und Maria erstellen einen Boxplot in TinkerPlots.</i>
<i>TinkerPlots-Kompetenz</i> mittel	Wenn Lernende einen konkreten Plan formulieren, diesen aber nur mit einem “trial-and-error-approach” umsetzen können, so beschreiben wir die <i>TinkerPlots-Kompetenz</i> als mittel	Hilde und Iris sind sich unsicher welchen Schalter sie in TinkerPlots nutzen müssen, um sich das arithmetische Mittel anzeigen zu lassen.
<i>TinkerPlots-Kompetenz</i> niedrig	Wenn Lernende einen konkreten Plan formulieren, diesen aber nicht umsetzen können, so beschreiben wir die Kompetenz mit <i>TinkerPlots-Kompetenz</i> niedrig.	Laura & Ricarda wollen einen Boxplot in TinkerPlots erstellen. Nach einiger Zeit sagt Laura: “Ich weiß nicht wie es geht”

Tabelle 6: Ausschnitt aus dem Kodiermanual zur Einschätzungsdimension „Verteilungsvergleichskompetenz“

Ausprägung	Hohe Kompetenz	Mittlere Kompetenz	Niedrige Kompetenz
Zentrum	Unterschiede der Lagemaße (aMittel/Median) werden quantifiziert und interpretiert.	Unterschiede der Lagemaße (aMittel/Median) werden qualitativ herausgearbeitet und nicht interpretiert.	Unterschiede der Lagemaße (aMittel/Median) werden falsch herausgearbeitet.
Streuung	Streuungsmaße (IQR) und informelle Angaben der Streuung (“dichter”, etc.) werden verglichen und interpretiert.	Streuungsmaße (IQR) und informelle Angaben der Streuung (“dichter”, etc.) werden verglichen und nicht interpretiert.	Streuungsunterschiede werden falsch verglichen.
Verschiebung	Die Verschiebung zwischen zwei Verteilungen ist korrekt herausgearbeitet (entweder durch Vergleich der mittleren 50 % der Verteilungen oder anhand des Vergleichs nicht-korrespondierender Kennzahlen)	Die Verschiebung zwischen den beiden Verteilungen ist qualitativ beschrieben.	Die Verschiebung zwischen beiden Verteilungen ist falsch herausgearbeitet.
Form	Form beider Verteilungen ist korrekt beschrieben und die Unterschiede zwischen den Verteilungen sind korrekt interpretiert.	Form beider Verteilungen ist korrekt beschrieben, aber die Unterschiede wurden nicht interpretiert.	Unterschiede zwischen der Form beider Verteilungen werden falsch herausgearbeitet.
p-basiert	p-basierte Unterschiede werden identifiziert und interpretiert.	p-basierte Unterschiede werden identifiziert aber nicht interpretiert.	P-basierte Unterschiede werden falsch herausgearbeitet.
q-basiert	q-basierte Unterschiede werden identifiziert und interpretiert.	q-basierte Unterschiede werden identifiziert aber nicht interpretiert.	q-basierte Unterschiede werden falsch herausgearbeitet.

Tabelle 7: Beispiele aus dem Kodiermanual zur Einschätzungsdimension „Verteilungsvergleichskompetenz“

Ausprägung	Hohe Kompetenz	Mittlere Kompetenz	Niedrige Kompetenz
Zentrum	“Die Arbeitnehmer verdienen durchschnittlich 29,5 % mehr als die Arbeitnehmerinnen” (Laura & Ricarda)	“Das arithmetische Mittel ist bei den Arbeitnehmern höher als bei den Arbeitnehmerinnen” (Hilde & Iris)	Kein Beispiel.
Streuung	“Die mittleren 50 % streuen bei den Männern mehr als bei den Frauen” (Hilde & Iris)	Die Interquartilsabstände der Verteilungen sind fast identisch. (Conrad & Maria)	Kein Beispiel.
Verschiebung	Der Median ist bei den Männern fast gleich dem ersten Quartil der Frauen (kein Beispiel aus den Daten)	“Die mittleren 50 % der Männer liegen weiter rechts als die der Frauen” (Hilde & Iris)	Kein Beispiel.
Form	„Die Verteilung der Männer hat einige Gipfel, aber die Verteilung scheint mehr rechtschief zu sein. So wird es mehr Arbeitnehmerinnen geben die weniger verdienen als Arbeitnehmer.“ (Laura & Ricarda)	„Hier [Verteilung des Einkommens bei den Arbeitnehmerinnen] können wir einen Gipfel bei 400€ finden. Bei den Männern gibt es auch ein Gipfel, aber der ist nicht so hoch.“ (Laura & Ricarda)	Kein Beispiel.
p-basiert	10 % der männlichen Arbeitnehmer verdienen mehr als 5000€ und 2 % der weiblichen Arbeitnehmer verdienen mehr als 5000€, also gibt es in der Gruppe der männlichen Arbeitnehmer mehr Vielverdiener als in der Gruppe der weiblichen Arbeitnehmer. (kein Beispiel aus den Daten)	„10 % der Arbeitnehmer verdienen mehr als 5000€, nur 2 % der Arbeitnehmerinnen verdienen mehr als 5000€.“ (Sandra & Luzie)	Kein Beispiel.
q-basiert	Das erste Quartil der Verteilung der männlichen Arbeitnehmer ist größer als der erste Quartil der Verteilung der weiblichen Arbeitnehmer, also verdienen die unteren 25 % der männlichen Arbeitnehmer tendenziell mehr als die unteren 25 % der weiblichen Arbeitnehmer.	Die unteren 25 % der Arbeitnehmerinnen sind geringer als die unteren 25 % der Arbeitnehmer.“ (Laura & Ricarda)	Kein Beispiel.