

2 Grundlegendes zum Themenkomplex „Datenanalyse“

Zu Beginn soll in diesem Abschnitt reflektiert werden, was „Statistisch denken und forschen lernen“ in dieser Arbeit bedeuten soll. Es werden einige Aspekte der sogenannten Datenkompetenz beleuchtet und dabei wird ein Blick darauf geworfen, was seitens der Bildungsstandards Mathematik von Schülern in der Primar- und Sekundarstufe und seitens der Vorgabe von Fachverbänden von angehenden Lehrern in diesem Inhaltsbereich verlangt wird. Außerdem werden in diesem Kapitel grundlegende Ideen der Umsetzung und Vermittlung einer Datenkompetenz (für Schule und Hochschule) vorgestellt und reflektiert.

2.1 Datenkompetenz

2.1.1 Allgemeine Datenkompetenz

„Daten sind überall“⁷, denn Diagramme und Tabellen begegnen uns täglich in den Medien. Die Interpretation dieser Daten beeinflusst unser Leben maßgeblich, wie beispielsweise politische oder sozialwissenschaftliche Entscheidungen, die auf Grundlage von Daten getroffen werden. (siehe u.a. Krüger 2012a)

In vielen Fällen lassen sich Manipulationen von Graphiken vornehmen, um bestimmte Positionen zu vertreten und unerwünschte Phänomene zu verschleiern, beziehungsweise erwünschte Muster herauszuheben. (siehe u.a. Krämer 2003) Um sich in der Welt der Daten zu orientieren und diese richtig verstehen zu können, bedarf es einer gewissen Kompetenz, der so genannten „Datenkompetenz“. Wir wollen in dieser Arbeit die Aspekte des Kompetenzbegriffs nach Weinert (2001) verfolgen. Franz Weinert (2001) definiert Kompetenz als

„die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ (Weinert 2001, 27).

Biehler und Weber (1995, 5) betonen, dass „Datenkompetenz [...] ähnlich wie Medienkompetenz oder soziale Kompetenz schwierig zu beschreiben [ist].“ Sie führen weiter aus, dass man von "Kompetenz" immer dann spricht, wenn man vor dem Dilemma steht, eine ganze Fülle von Fertigkeiten, Fähigkeiten, Begriffen, Einsichten, etc. zu meinen, sich aber kaum in der Lage sieht, diese umfassend zu beschreiben. (vgl. Biehler und Weber 1995, 5)

7 Einen eindrucksvollen Beleg dazu liefert Gould (2011) in seiner Keynote zum Auftakt der SRTL-7 Tagung (July 2011, Texel, The Netherlands).

Hancock (1995) und Biehler (2001) erörtern, dass unter Datenkompetenz ebenfalls zahlreiche Teilfertigkeiten wie Erheben, Strukturieren, Darstellen und Interpretieren von Daten zu verstehen sind. Wagner (2006) hat im Rahmen ihrer Staatsexamensarbeit Datenkompetenz in die folgenden neun Punkte untergliedert (siehe auch Biehler 2001, 98):

- (1) Organisation von Daten, Datenstrukturen, Variablentypen
- (2) Graphiken und Diagramme zur Datenanalyse und ihre Reichweite
- (3) Statistische Begriffe (Mittelwerte, Streuungsmaße, usw....)
- (4) Planung einer Analyse, Problemformulierung
- (5) Methodenauswahl und -reflexion
- (6) Interpretation von Darstellungen und Ergebnissen
- (7) Maßstäbe für akzeptable Begründungen und Schlussfolgerungen
- (8) Kommunikation von Resultaten einer Analyse
- (9) Umgang mit einem Softwarewerkzeug zur Datenanalyse

Wir sehen, der Begriff der Datenkompetenz umfasst verschiedene Facetten und wir können verschiedene Anforderungen unterscheiden. Zentral, mit Blick auf die Anforderungen im täglichen Leben, ist sicherlich das Lesen und Interpretieren von Graphiken in den Medien (z.B. Diagramme in der Zeitung), für die es eine „Graphikkompetenz“ bedarf.

2.1.1.1 *Graphikkompetenz*

Die oben angesprochene Graphikkompetenz könnte man, nach dem Vorbild von Friel, Curcio und Bright (2001), in drei Ebenen des Verständnisses von graphischen Darstellungen unterteilen: „reading the data“, „reading between the data“ und „reading beyond the data“. Diese drei Ebenen finden sich auch bei der von Gonzalez, Espinel und Ainley (2011, 190) definierten Graphikkompetenz („Graphical competence“) wieder. Gonzalez, Espinel und Ainley (2011, 190) verstehen unter „Graphical competence“ die Vereinigung der folgenden drei Aspekte:

- „The ability to extract data from different sorts of graphs and to interpret meanings from them by reading between, beyond, and behind the data displayed to form hypotheses about the phenomena represented in the graph
- The capacity to select and create appropriate graphs for specific situations, with or without the support of technology; and
- The ability to critically evaluate graphs and to distinguish the relative strengths and limitations of particular graphical representations, recognizing that creating a graph involves an interpretation of the original data.“ (Gonzales, Espinel und Ainley 2011, 190)

Eine Übersicht über den Stand der Forschung rund um die „Graphikkompetenz“ von Schülern findet sich in Gonzalez, Espinel und Ainley (2011) und soll an dieser Stelle nicht weiter ausgeführt werden. Wir wollen uns in dieser Arbeit an der Definition der Graphikkompetenz von Gonzalez, Espinel und Ainley (2011, 190) orientieren.

2.1.2 Statistical Literacy, Statistical Reasoning und Statistical Thinking

Chris Hancock (1995, 34) prägt den Begriff [Statistical] „Literacy“ als umfassendes System von Praktiken und Begriffen, das erst über einem längeren Zeitraum entwickelt werden kann. Im Allgemeinen meint er eine Vielzahl von Fertigkeiten, Fähigkeiten, Begriffen, Einsichten, etc., die sich über einen längeren Zeitraum hinweg entwickeln und welche notwendig oder zumindest hilfreich beim sachgerechten Umgang mit Daten sind. Statistical Literacy umfasst Fähigkeiten, die benötigt werden, um statistische Informationen zu verstehen. Die Wichtigkeit von „Statistical Literacy“ betonen Konold und Higgins (2003, 193):

„At the practical level, knowledge of statistics is a fundamental tool in many careers, and without an understanding of how samples are taken and how data are analyzed and communicated, one cannot effectively participate in most of today’s important political debates about the environment, health care, quality of education, and equity. For those who have traditionally been left out of the political process, probably no skill is more important to acquire in the battle for equity than statistical literacy.“ (Konold und Higgins 2003, 193)

Vermehrt tritt die Position auf, die Definition der „Statistical Literacy“ zu präzisieren. Definitionen und Abgrenzungen der drei Bereiche „Statistical Literacy“, „Statistical Thinking“ and „Statistical Reasoning“ finden sich in Ben-Zvi und Garfield (2004, 7):

„Statistical literacy includes the skills that might be used to understand statistical information or research results. Statistical reasoning is the way in which people reason with statistical ideas and make sense out of statistical information. Statistical thinking involves an understanding of why and how statistical investigations are conducted and the “big ideas” that underlie statistical investigations.“ (Ben-Zvi und Garfield 2004, 7)

Die höchste Stufe, die es in diesem Schema zu erlangen gilt, ist „Statistical Thinking“. Hier entwickelt sich ein Verständnis für das Durchführen von statistischen Untersuchungen und zum anderen für die sogenannten „Big ideas“, die von Moore (1990, 135) im Sinne von „fünf Kernelementen statistischen Denkens“ postuliert worden sind:

- “The omnipresence of variation in processes
 - The need for data about processes
 - The design of data production with variation in mind
 - The quantification of variation
 - The explanation of variation”
- (Moore 1990, 135)

Besonders die Berücksichtigung von Variation spielt in den „Big ideas“ von Moore eine große Rolle. Er spricht von einer Omnipräsenz der Variation („omnipresence of variation in processes“) in den Daten und regt an, Variation zu quantifizieren („the quantification of variation“) und zu erklären („the explanation of variation“). Ebenfalls sieht er die Erhebung von Daten aus Prozessen („the need for data about processes“) und das Design der Erhebung unter dem Bewusstsein der Variation als wichtige Idee an („the design of data production with variation in mind“). Pfannkuch und Wild (2004, 19) regten

knapp 14 Jahre später die „fünf Typen statistischen Denkens“ an, welche sich auch in ihrem propagierten PPDAC-Zyklus (Wild und Pfannkuch 1999) wiederfinden (siehe Kapitel 2.4.1):

- “Recognition of the need for data
 - Transnumeration
 - Consideraton of variation
 - Reasoning with statistical models
 - Integrating the statistical and contextual”
- (Pfannkuch und Wild 2004, 19)

Wie schon Moore (1990) fordern auch Wild und Pfannkuch (1999) und Pfannkuch und Wild (2004), dass Lernende das Bedürfnis der Erhebung von Daten („need for data“) erkennen sollen. Bei der Arbeit mit Daten sehen sie „Transnumeration“, das Wahrnehmen von Variation, sowie das Denken in stochastischen Modellen als fundamental an. „Transnumeration“ ist die Fähigkeit des Wechsels von Darstellungen, um bestimmte Muster in den Daten ausfindig zu machen. Für eine Definition siehe Pfannkuch und Wild (2004) oder Shaughnessy (2007, 963). Auch die Integration beider Welten „Kontext“ und „Statistik“ ist laut Wild und Pfannkuch (1999) von Bedeutung.

Rossman, Chance und Lock (2001, 48) formulieren „Leitideen“ explizit für die Interpretation von Verteilungen. Zu den einzelnen Komponenten (Zentrum einer Verteilung, Streuung einer Verteilung, etc.) stellen sie Leitfragen, welche sich auch für den Vergleich von Verteilungen interpretieren lassen. Sie nennen hier das Zentrum, die Streuung, sowie die Form einer Verteilung. Ebenso weisen sie auf die Berücksichtigung von Teilgruppen, Ausreißern und Strukturen beim Beschreiben und Interpretieren von Verteilungen hin.

- „Center of distribution“
- „Distribution’s variability“
- „Shape of a distribution“
- “A distribution may have peaks or clusters”
- “Outliers”

(entnommen aus: Rossman et al. 2001, 48)

Zu diesen Aspekten könnte man abschließend folgende Fragen zur Verteilung formulieren:

- Zentrum („center of distribution“): Wo liegen die Daten?
- Streuung („Distribution’s variability“): Wie streuen die Daten?
- Form / Gestalt der Verteilung („Shape of distribution“): Ist die Verteilung symmetrisch, linkssteil/rechtsschief, rechtssteil/linksschief?
- Teilgruppen („Peaks & Cluster“): Gibt es Hinweise auf Teilgruppen?
- Ausreißer („Outliers“): Gibt es Daten, die sich auffällig vom Muster in den Daten abheben?

Diese Leitfragen bieten eine gute Grundlage, um eine Verteilung eines numerischen Merkmals zu interpretieren.

2.2 Anforderungen an Schüler, Lehramtsstudierende und Lehrer im Bereich der deskriptiven Statistik

Wir haben in 2.1 erste Eindrücke erhalten, was Datenkompetenz ausmacht, was darunter zu verstehen ist und in welche verschiedenen Facetten wir diese gliedern können. Es stellt sich die Frage, welche dieser Komponenten für die Schule und somit auch für die Lehramtsanwärter im Bereich des Grund-, Haupt-, Real- und Gesamtschullehramts von Bedeutung sind. Dazu betrachten wir in diesem Abschnitt inhaltsbezogene Anforderungen im Bereich der deskriptiven Statistik: zum einen an Schüler der Primarstufe und der Sekundarstufe I, wie sie in den Bildungsstandards sowie in Empfehlungen einschlägiger Verbände formuliert werden und zum anderen Anforderungen an Lehrkräfte in Mathematik für Grund-, Haupt-, Real- und Gesamtschulen. Insgesamt betrachten wir dazu die Bildungsstandards für Grundschule Mathematik (Walther, Van den Heuvel-Panhuizen, Granzer und Köller 2012), die Bildungsstandards Mathematik für die Sekundarstufe I (Blum et al. 2006) sowie die Empfehlungen des Arbeitskreises Stochastik der Gesellschaft für Didaktik der Mathematik zum Abschlussniveau der Schüler nach der Sekundarstufe⁸. Im Rahmen der Anforderungsbeschreibung der Lehrkräfte beziehen wir uns auf Empfehlungen des Arbeitskreises Stochastik für die Ausbildung der Primarstufen-Lehramtsstudierenden sowie auf die Empfehlungen von GDM, DMV und MNU zur Lehrerbildung für den Bereich „Beschreibende Statistik/Datenanalyse“.

2.2.1 Anforderungen an Schüler im Bereich der deskriptiven Statistik

Zunächst betrachten wir Anforderungen an Schüler der Primarstufe im Bereich der deskriptiven Statistik im Sinne der Bildungsstandards Mathematik für die Primarstufe. Ein Kern der stochastischen Ausbildung der Schüler umfasst die Ausbildung dieser zum „mündigen Bürger“⁹. Das beginnt bereits mit den formulierten Leitideen zu „Daten, Häufigkeit und Wahrscheinlichkeit“ in der Primarstufe (siehe Hasemann und Mirwald 2012). Dort wird gefordert, dass die Schüler lernen sollen...

- „...wie man Daten über Objekte oder Ereignisse erfasst.
- ...wie man sie dokumentiert, insbesondere dann, wenn sie flüchtig (vergänglich) sind.
- ...dass es erforderlich ist, vor der Datenerhebung Kriterien oder Merkmale festzulegen, nach denen die beobachteten Objekte oder Ereignisse unterschieden werden sollen.
- ...wie man die so erfassten Daten für andere Personen übersichtlich in Tabellen und Diagrammen darstellt.
- ...dass es hilfreich oder sogar notwendig sein kann, die Daten noch weiter zu bearbeiten um ihren Informationsgehalt zu erhöhen.

8 <http://www.mathematik.uni-dortmund.de/ak-stoch/stellung.html> (aufgerufen am 18.5.2015)

9 Siehe KMK-Bildungsstandards von 2004.

- ...wie man solchen Darstellungen Informationen entnimmt und diese dann benutzt.“ (Hasemann und Mirwald 2012, 145)

Die Empfehlungen des GDM-Arbeitskreises Stochastik fordern unter anderem, dass Schüler am Ende der Primarstufe „Probleme kennen und Fragen selbst stellen können, die sich mit Hilfe von Daten beantworten lassen“, „erste Erfahrungen im Erfassen und Aufbereiten von Daten mit Strichlisten, Häufigkeitstabellen, Strecken- und Streifendiagrammen besitzen“ und „Informationen aus einfachen Diagrammen entnehmen können“. ¹⁰ Betrachtet man die Formulierungen der Leitideen „Daten, Häufigkeit und Wahrscheinlichkeit“, sowie die Empfehlungen des GDM-Arbeitskreises Stochastik, so sehen wir, dass Schüler schon auf einer frühen Stufe (Primarstufe) mit dem Problem der Datenerhebung und der Dokumentation, Strukturierung und Interpretation von Daten konfrontiert werden sollten. Ideen für die Umsetzung und für eine frühe Förderung von Datenkompetenz in der Primarstufe finden sich in Biehler und Frischemeier (2013) und Biehler und Frischemeier (2015a).

An die Leitidee „Daten, Häufigkeit und Wahrscheinlichkeit“ anknüpfend, sieht die Leitidee „Daten und Zufall“ im Rahmen der Bildungsstandards Mathematik für den mittleren Schulabschluss (KMK 2004) für die weiterführenden Schulformen unter anderem vor, dass

„Schüler graphische Darstellungen und Tabellen von statistischen Erhebungen auswerten, statistische Erhebungen planen, Daten systematisch sammeln, in Tabellen erfassen und sie graphisch – auch unter Verwendung geeigneter Hilfsmittel, beispielsweise Software – darstellen, Daten unter der Verwendung von Kenngrößen interpretieren und Argumente, die auf einer Datenanalyse basieren, reflektieren und bewerten.“ ¹¹ (KMK 2004)

Ausführlich findet man dort in der Leitidee „Daten & Zufall“ (Biehler und Hartung 2006) die folgenden Aspekte:

„Die Schülerinnen und Schüler

- werten graphische Darstellungen und Tabellen von statistischen Erhebungen aus,
 - planen statistische Erhebungen,
 - sammeln systematisch Daten, erfassen sie in Tabellen und stellen sie graphisch dar, auch unter Verwendung geeigneter Hilfsmittel (wie Software),
 - interpretieren Daten unter Verwendung von Kenngrößen,
 - reflektieren und bewerten Argumente, die auf einer Datenanalyse basieren,
 - beschreiben Zufallserscheinungen in alltäglichen Situationen,
 - bestimmen Wahrscheinlichkeiten bei Zufallsexperimenten.“
- (Biehler und Hartung 2006, 52)

Vogel und Eichler (2010, 879) fassen diese so zusammen,

10 URL: <http://www.mathematik.uni-dortmund.de/ak-stoch/stellung.html> (aufgerufen am 10.11.2014)

11 Siehe KMK-Bildungsstandards von 2004.

„dass es grundsätzlich darum geht, dass die Schülerinnen und Schüler lernen, Fragen an empirische Phänomene ihrer erlebten Umwelt zu stellen und mit elementaren mathematischen Mitteln der Sekundarstufe I zu beantworten. Die Daten sind der Wahrscheinlichkeit vorgeordnet und der statistische Aspekt geht über das bloße Erstellen von Graphiken als Teil des Sachrechnens deutlich hinaus.“ (Vogel und Eichler 2010, 879)

Der GDM-Arbeitskreis Stochastik (2003) hat ebenfalls eine Empfehlung herausgegeben, welches Abschlussniveau Schüler bezüglich ihrer stochastischen Kompetenzen am Ende der Sekundarstufe I haben sollten. So sollen diese unter anderem „Strichlisten und Häufigkeitstabellen für eindimensionale Daten anfertigen, sowie relative Häufigkeiten berechnen“ können. Außerdem sollen sie

„verschiedene Möglichkeiten zur graphischen Darstellung (wie Kreisdiagramm, usw.) kennen, und in der Lage sein, angemessene grafische Darstellungen für Daten auszuwählen und in einfachen Fällen zu erstellen, wobei nach Möglichkeit geeignete Software eingesetzt werden sollte. Weiterhin sollten [im Realschul- und gymnasialen Bildungsgang] auch Boxplots verwendet [und] mehrere Verteilungen mittels Boxplots miteinander verglichen [...] werden.“ (Empfehlungen GDM-Arbeitskreis Stochastik 2003, 23)

Ebenso wird gefordert, dass Schüler

„vorliegende graphische Darstellungen lesen und interpretieren können [...], das arithmetische Mittel und den Zentralwert (Median) bestimmen, interpretieren und dessen angemessene Verwendung beurteilen können, sowie qualitativ das Problem der Streuung verstehen und ein einfaches Streuungsmaß (Spannweite) berechnen und interpretieren können. Darüber hinaus sollten Schülerinnen und Schüler im Realschul- oder gymnasialen Bildungsgang ein weiteres Streuungsmaß (z.B. die mittlere Abweichung, die Vierteldifferenz und oder die Standardabweichung) an Beispielen berechnen und interpretieren können.“ (Empfehlungen GDM-Arbeitskreis Stochastik 2003, 24)

Im Realschul- und gymnasialen Bildungsgang wird ferner gefordert, dass „die Schülerinnen und Schüler das Problem der Gruppierung der Daten kennen und in einfachen Fällen eine Klassenbildung vornehmen und dafür das arithmetische Mittel berechnen können.“ Als weitere Verfeinerung wird für Schüler des gymnasialen Bildungsgangs gefordert, dass diese „Histogramme erstellen können und über ihre grundlegenden Eigenschaften wissen.“ Schließlich sollen die Schüler „auf Grundlage von Daten Schlussfolgerungen und Prognosen qualitativ herleiten und bewerten können und [...] insbesondere Unterschiede zwischen den Ergebnissen verschiedener Stichproben einer Grundgesamtheit untersuchen.“ Hieran sollen sie „begründete Vermutungen aufstellen, neue Fragen formulieren und entsprechende neue Untersuchungen planen.“

Zusammenfassend lässt sich sagen, dass von den Schülern ein breites Anforderungsspektrum (Datenerhebung, Durchführung, Dokumentation, Analyse, etc.) hinsichtlich der Datenanalyse in Primar- und Sekundarstufe gefordert wird. Was von Schülern verlangt wird, wird auch von Lehrern in besonderem Maße gefordert. Während bei den Schülern der fachliche Aspekt im Vordergrund steht, muss man in der Lehrerbildung auch die entsprechende Fachdidaktik sowie ggfs. den Einsatz von Technologien miteinbeziehen. So scheint es sinnvoll, dass angehende Lehrkräfte selbst Aufgaben rund um

den Themenkomplex Datenanalyse (mit Software) bearbeiten und mögliche Schwierigkeiten dabei erfahren. Schließlich müssen Lehrer, wenn sie Datenanalyse unterrichten wollen, diese Thematik vorher selbst erlernt und mögliche Schwierigkeiten innerhalb der Lernprozesse erfahren haben, um Fehler im Lernprozess der Schüler aufdecken zu können und bei Problemen angemessene Hilfestellungen leisten zu können.

2.2.2 Anforderungen an Lehramtsstudierende und Lehrer im Bereich der deskriptiven Statistik

Die Empfehlungen vom Arbeitskreis Stochastik (2012) sehen für die Ausbildung der Primarstufen-Lehramtsstudierenden den Erwerb von Kompetenzen in den folgenden drei Aspekten vor:¹²

- Kompetenzen im Erkennen und Analysieren von Erscheinungen mit Zufallscharakter
- Kompetenzen in der Planung, Durchführung und Auswertung statistischer Untersuchungen
- Kompetenzen in der Ermittlung und Interpretation von Wahrscheinlichkeiten

In dieser Arbeit ist vor allem der mittlere Aspekt „Kompetenzen in der Planung, Durchführung und Auswertung statistischer Untersuchungen“ von Bedeutung. Hier wird von den Studierenden des Lehramts Mathematik an der Primarstufe Folgendes gefordert:¹³

„Die Studierenden/die Lehrkräfte

- können Fragen stellen, die sich mit Hilfe von statistischen Untersuchungen beantworten lassen,
- beherrschen grundlegende Vorgehensweisen bei der Planung einer statistischen Untersuchung,
 - insbesondere kennen sie
 - Probleme der Auswahl einer Stichprobe und können eine solche in einfachen Fällen durch zufällige Auswahl gewinnen,
 - ausgewählte Probleme der Erstellung von Fragen und können zu einfachen Sachverhalten geeignete Fragen entwickeln,
 - exemplarisch mögliche Fehler bei der Planung von statistischen Untersuchungen,
- können sicher Strichlisten und Häufigkeitstabellen für eindimensionale Daten anfertigen sowie relative Häufigkeiten berechnen,
- kennen sicher folgende Möglichkeiten zur grafischen Darstellung von eindimensionalen Daten: Kreisdiagramm, Streckendiagramm (Stabdiagramm), Streifendiagramm (Balken- oder Säulendiagramm), Liniendiagramm (Kurvendiagramm, Streckenzug, Polygonzug) und Bilddiagramm (Piktogramm),
- können angemessene grafische Darstellungen für Daten auswählen und erstellen, wobei sie auch geeignete Software verwenden,
- können vorliegende grafische Darstellungen lesen und interpretieren,

12 http://www.mathematik.uni-dortmund.de/ak-stoch/Empfehlungen_Stochastik_Grundschole.pdf (aufgerufen am 19.9.2014)

13 Siehe ebenfalls: http://www.mathematik.uni-dortmund.de/ak-stoch/Empfehlungen_Stochastik_Grundschole.pdf (aufgerufen am 19.9.2014)

- kennen fehlerhafte grafische Darstellungen
- können sicher das arithmetische Mittel einer Häufigkeitsverteilung bestimmen, interpretieren und dessen angemessene Verwendung beurteilen,
- verstehen qualitativ das Problem der Streuung, können sicher die Spannweite interpretieren und kennen exemplarisch weitere Streuungsmaße,
- kennen folgende Mittel und Methoden der Explorativen Datenanalyse und können sie in geeigneten einfachen Fällen sicher anwenden und mit den Mitteln und Methoden der klassischen beschreibenden Statistik vergleichen: Stamm-Blätter-Diagramm (Stängel-Blätter-Diagramm), Boxplot, Zentralwert (Median), Viertelwerte (Quartile), Vierteldifferenz,
- kennen exemplarisch Probleme der Gruppierung von Daten und können in einfachen Fällen eine Klassenbildung vornehmen, das arithmetische Mittel näherungsweise berechnen und Histogramme erstellen,
- können auf der Grundlage von Daten Schlussfolgerungen und Prognosen qualitativ herleiten und bewerten, insbesondere nach Beziehungen zwischen der Ausprägung der Bedingungen und der Verteilung der Daten suchen, begründete Vermutungen aufstellen, neue Fragen formulieren und dazu entsprechende neue Untersuchungen planen.“
(Empfehlungen vom Arbeitskreis Stochastik 2012)

Hier zeigt sich ein breites Anforderungsspektrum an zukünftige Mathematik-Lehrer der Primarstufe. Neben dem selbstständigen Durchführen einer Datenerhebung (mit Formulierung adäquater Fragen, Erstellen eines Erhebungsinstruments und eigenständiger Erhebung der Daten), wird auch das Kennen und Anwenden vielfältiger statistischer Konzepte und Darstellungen beim Prozess der Datenanalyse verlangt. Ebenso werden die Kenntnis geeigneter Software sowie die Interpretation grafischer Darstellungen gefordert. Auffällig ist, dass in diesen Forderungen auch explizit das „Kennen von Fehlern“ bzw. „Kennen von Problemen“ formuliert ist, die Lehramtsstudierende sicherlich erfahren müssen, um in ihrem späteren Beruf mit den Schwierigkeiten der Schüler im Lernprozess angemessen umgehen zu können.

In den Empfehlungen von GDM, DMV und MNU (2008)¹⁴ zur Lehrerbildung finden sich ebenfalls Kompetenzempfehlungen für den Bereich „Beschreibende Statistik/Datenanalyse“. Dort wird gefordert, dass...

- „die Studierenden statistische Erhebungen (Befragung, Beobachtung oder Experiment) planen, durchführen und auswerten
- die Studierenden grafische Darstellungen für uni- und bivariate Daten (z.B. Kreuztabelle) lesen und erstellen und deren Eignung für die jeweilige Fragestellung bewerten
- die Studierenden uni- und bivariate Kennwerte (z.B. Mittelwerte, Streumaße, Korrelationen, Indexwerte) bestimmen, verwenden und diese angemessen interpretieren.“ (Empfehlungen von GDM, DMV und MNU 2008)

14 http://madipedia.de/images/2/21/Standards_Lehrerbildung_Mathematik.pdf
(aufgerufen am 19.9.2014)

Unter einem weiteren Punkt „Neue Medien“ fordern sie u.a. die Verwendung von Tabellenkalkulationsprogrammen sowie die Verwendung von statistischer Software zur Darstellung und explorativen Analyse von Daten. Die Empfehlungen von GDM, DMV und MNU sind zwar knapp gehalten, enthalten im Kern aber die Aspekte, die auch vom AK Stochastik gefordert werden.

Abschließend fällt bei der Betrachtung der Standards (sei es für Schüler oder für Lehrer) auf, dass insbesondere die Kompetenzen bzgl. des Planens einer eigenen Datenerhebung, des eigenen Erhebens von Daten, sowie des anschließenden Auswertens vielfach gefordert werden. Ebenso wird sowohl von Schülern als auch von Lehrern ein kompetenter Umgang mit statistischer Software verlangt. Dies sind Aspekte, die wir mit einer eigens konzipierten Lehrveranstaltung für Studierende des Lehramts Mathematik an Grund-, Haupt-, Real- und Gesamtschulen vermitteln wollen. Fachlich wollen wir uns dabei insbesondere auf das Vergleichen von Verteilungen eines numerischen Merkmals, welches wir als wichtige Teilkompetenz der Datenkompetenz ausmachen, beschränken.

2.3 „Denken in Verteilungen“ als fundamentaler Bestandteil einer Datenkompetenz

In diesem Kapitel sollen verschiedene normative Aspekte zur fundamentalen Idee der Verteilung in der Statistik ausgeführt werden. Dabei wird zunächst zwischen Verteilungen kategorialer und numerischer Merkmale unterschieden. Im weiteren Verlauf werden dann ausgehend von verschiedenen Charakteristika von Verteilungen numerischer Merkmale mögliche Vergleichsaspekte beim Vergleich von Verteilungen eines numerischen Merkmals aufgezeigt.

2.3.1 Verteilungen kategorialer und numerischer Merkmale

Ein grundlegender Artikel zum Konzept der Verteilung ist der Artikel von Wild (2006). Wild (2006) sieht die Verteilung als Konzept, eine theoretische Perspektive aus der man Variation in den Daten betrachtet. Dabei benutzt er die Metapher der „Verteilung als Linse“ (Wild 2006, 11), die helfen soll, die Variation in den Daten zu erkennen. Dieses verdeutlicht die Abbildung 1. Die Variation in der realen Welt spiegelt sich als Variation in den Daten wieder. Die Verteilung ist dabei wie eine Linse mit der man auf die Verteilung schaut. Weiter führt Wild (2006, 11) aus:

„All of the information about patterns of variation is in the (typical multivariate) frequency distribution. All summary statistics and almost all the graphs we look at are summaries and graphs of frequency distributions. We use them to discover and describe aspects of patterns in the variation contained in the frequency distributions.“ (Wild 2006, 11)

Insbesondere unterscheidet Wild (2006) Stichprobenverteilungen, empirische Verteilungen und theoretische Verteilungen.

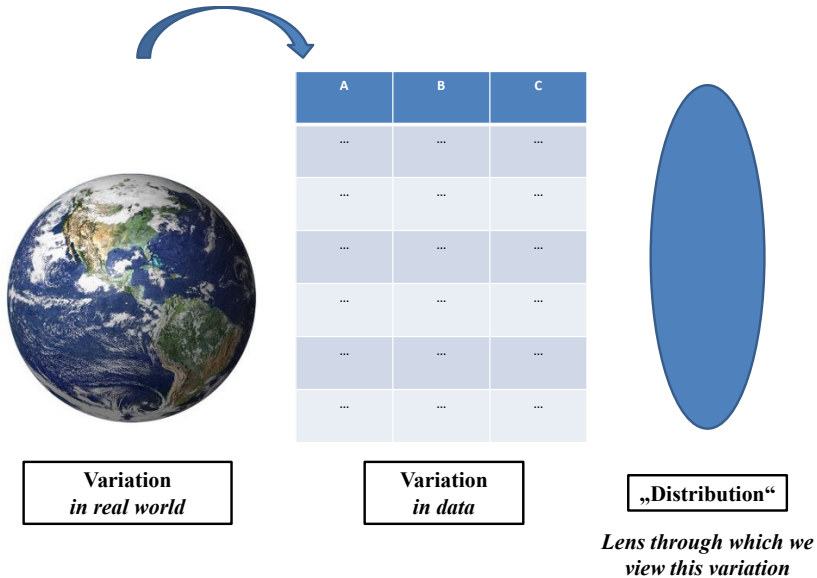


Abbildung 1: „Verteilung als Linse“ - Abbildung angelehnt an Wild (2006, 11)

Die Unterscheidung zwischen theoretischen und empirischen Verteilungen findet sich in Wild (2006, 13), die Unterscheidung zwischen Stichproben- und Populationsverteilung in Wild (2006, 18). Eine Unterscheidung zwischen theoretischen und empirischen Verteilungen nimmt auch Biehler (2007b, 3) vor und unterscheidet hier mehrere Facetten, zum einen in der beschreibenden Statistik, zum anderen in der Wahrscheinlichkeitstheorie:

„In der Beschreibenden Statistik geht es um die (empirische) Verteilung der Ausprägungen eines oder mehrerer Merkmale in einer Stichprobe, in der Wahrscheinlichkeitstheorie geht es um (theoretische) Wahrscheinlichkeitsverteilungen. Bei Vergrößerung des Stichprobenumfangs nähern sich unter bestimmten Bedingungen die empirischen Häufigkeitsverteilungen immer mehr der theoretischen Wahrscheinlichkeitsverteilung an (Gesetz der großen Zahl). In der beurteilenden Statistik schließt man von empirischen Verteilungen auf Merkmale theoretischer Verteilungen.“ (Biehler 2007b, 3)

Ben-Zvi und Garfield (2004, 400) sehen in einer Verteilung eine Darstellung quantitativer (numerischer) Merkmale, die anhand verschiedener Eigenschaften („shape, center, and spread, [...]“) beschrieben werden kann:

„a representation of quantitative data that can be examined and described in terms of shape, center, and spread, as well as unique features such as gaps, clusters, outliers, and so on.“ (Ben-Zvi und Garfield 2004, 400)

Diese Arbeit beschäftigt sich vor allem mit „empirischen Verteilungen der Ausprägungen eines oder mehrerer Merkmale in einer Stichprobe“ und greift daher die „empirische Häufigkeitsverteilung“ heraus. Unter Häufigkeitsverteilung versteht man im Allgemeinen die Zuordnung von Häufigkeiten zu Merkmalsausprägungen:

„Die Zuordnung von Häufigkeiten zu den Merkmalsausprägungen heißt Häufigkeitsverteilung. Die Maßzahlen in den vorangehenden Abschnitten kennzeichnen ausgewählte Eigenschaften (speziell die mittlere Lage und die Variation betreffend) einer Häufigkeitsverteilung von Messwerten.“ (Sachs und Hedderich 2006, 80)

Nach Fahrmeir, Künstler, Pigeot und Tutz (2007, 33ff.) gibt es mehrere Möglichkeiten, Verteilungen zu beschreiben: mit graphischen Darstellungen und durch Kennzahlen. Dabei gibt es Unterschiede beim Beschreiben von Verteilungen kategorialer und numerischer Merkmale. So unterscheiden Fahrmeir et al. (2007, 19) zwischen qualitativen/kategorialen und quantitativen Merkmalen, wie folgt:

„Unter *qualitativen* oder *kategorialen* Merkmalen versteht man Größen, die endlich viele Ausprägungen besitzen und höchstens ordinalskaliert sind. Von Bedeutung ist dabei, dass die Ausprägungen eine Qualität und nicht ein Ausmaß widerspiegeln. Geben die Ausprägungen hingegen eine Intensität bzw. ein Ausmaß wieder, in dem die interessierende Eigenschaft enthalten ist, so spricht man von *quantitativen* Merkmalen. Damit sind alle Messungen im herkömmlichen Sinn, deren Werte Zahlen darstellen, Ausprägungen quantitativer Merkmale. Somit lässt sich auch direkt wieder ein Bezug herstellen zum Skalenniveau: Kardinalskalierte Merkmale sind stets ebenfalls quantitativ.“ (Fahrmeir et al. 2007, 19)

Fahrmeir et al. (2007, 19) führen außerdem aus:

„Bei ordinalskalierten Merkmalen ist die Zuordnung nicht so eindeutig. Sie nehmen eine Zwitterstellung ein. Da man ihre Ausprägungen anordnen kann, besitzen sie einen – wenn auch – schwachen quantitativen Aspekt. Allerdings ordnet man sie aufgrund ihres eher dominierenden qualitativen Charakters den qualitativen Merkmalen zu, zumindest wenn sie nur endlich viele Ausprägungen besitzen.“ (Fahrmeir et al. 2007, 19)

In dieser Arbeit werden die Merkmalstypen „nominalskalierte Merkmale“ sowie „ordinalskalierte Merkmale“ unter dem Sammelbegriff „kategoriales Merkmal“ und „kardinalskalierte“ sowie „intervallskalierte Merkmale“ unter dem Sammelbegriff „numerisches Merkmal“ gefasst. Eine Unterscheidung aller Merkmalstypen findet sich in Fahrmeir et al. (2007, 20).

Im Folgenden sollen zur Verdeutlichung der Unterscheidung von kategorialen und numerischen Merkmalen Graphiken für die Verteilungen entsprechender Merkmale vorgestellt werden. Die Graphiken wurden –bis auf eine Ausnahme– mit der Software Tin-

kerPlots erstellt. Eine genauere Erläuterung der Erstellung dieser Graphiken findet sich in Kapitel 3 dieser Arbeit.

Verteilung kategorialer Merkmale

Graphische Darstellungen, die Fahrmeir et al. (2007, 35f.) für die Veranschaulichung der Verteilung kategorialer Merkmale nennen, sind das Streifendiagramm, das Stabdiagramm¹⁵, das Säulendiagramm und das Kreisdiagramm. Als Beispiel zur Verteilung eines kategorialen Merkmals kann das Säulendiagramm in Abbildung 2 zur Verteilung des Merkmals „Monat_Geburtsstag“ (Geburtsmonat) aus den Rischenau-Daten (Dettmar 2013) betrachtet werden.

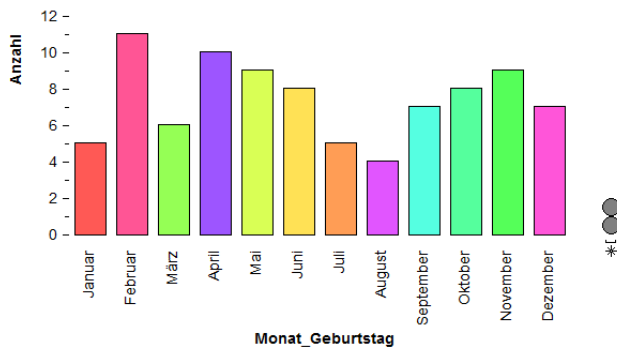


Abbildung 2: Verteilung des Merkmals „Geburtsmonat“ (Säulendiagramm) (Datensatz Dettmar 2013)

Jedem Monat (jeder Ausprägung) ist hier eine absolute Häufigkeit zugeordnet. Dabei ist die jeweilige Länge/Höhe der Säulen proportional zur Häufigkeit. Aus dieser Graphik kann man einige Aspekte entnehmen, wie z.B., dass die meisten Kinder im Februar Geburtstag haben oder, dass nur wenige Kinder in den Sommermonaten Juli und August Geburtstag haben. TinkerPlots bietet darüber hinaus verschiedene Farben für die verschiedenen Ausprägungen (Monate) an. Dieses ist im Allgemeinen nicht unbedingt üblich, oft sind die Säulen einfarbig gehalten. Eine weitere Darstellungsmöglichkeit die Verteilung eines kategorialen Merkmals darzustellen, die auch gerne schon im Mathematikunterricht der Grundschule aufgegriffen wird, ist das Kreisdiagramm. In Abbildung 3 ist die Verteilung des Merkmals „Wie_kommst_du_zur_Schule“ dargestellt. Man kann der Graphik entnehmen, dass ca. ein Viertel der befragten Schüler zu Fuß zur Schule und ca. 2/3 der befragten Schüler mit dem Bus zur Schule kommen.

15 Siehe auch Kütting (1994, 37ff.).

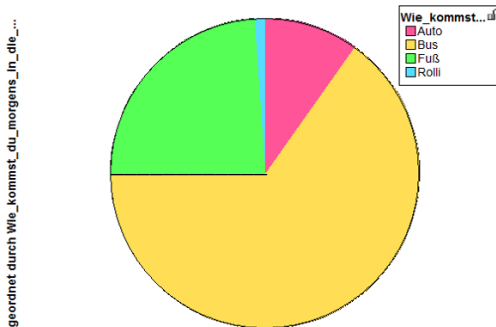


Abbildung 3: Verteilung des Merkmals „Wie_kommt_du_morgens_in_die_Schule?“ (Kreisdiagramm) (Datensatz Dettmar 2013)

Verteilung numerischer Merkmale

Eine Verteilung eines numerischen Merkmals kann u.a. -siehe Fahrmeir et al. (2007, 37ff.)- Durch ein gestapeltes Punktdiagramm, ein Histogramm, ein Boxplot oder auch durch ein Stängel-Blatt-Diagramm dargestellt werden. Abbildung 4 zeigt die Verteilung des Merkmals „Körpergröße“, auf der linken Seite dargestellt in einem gestapelten Punktdiagramm. Ebenfalls wäre es möglich, die Verteilung des Merkmals „Körpergröße“ in einem Histogramm (rechte Seite, Abb. 4) zu veranschaulichen.

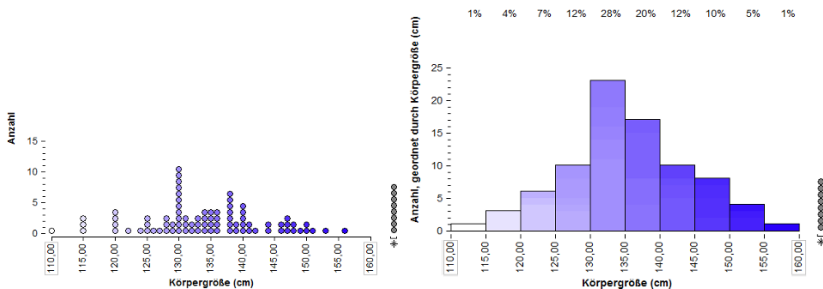


Abbildung 4: Verteilung des Merkmals „Körpergröße“ (gestapeltes Punktdiagramm, links) und Verteilung des Merkmals „Körpergröße“ (Histogramm, rechts) (Datensatz Dettmar 2013)

Im Punktdiagramm lassen sich im Allgemeinen Modalwert, Spannweite, die Form und eventuell auch das Zentrum einer Verteilung identifizieren. Ein Histogramm¹⁶ (wie in Abbildung 4, rechts) kann vertiefende Einsichten stiften, z.B. in die Lage des Zentrums

¹⁶ Nähere Informationen zur Definition des Histogramms, sowie verschiedene Typen von Histogrammen sind in Kütting (1994, 48-52) aufgeführt.

oder in die Form der Verteilung. Auch hier färbt TinkerPlots die Punkte (Abb. 4, links) bzw. Säulen (Abb. 4, rechts) ein - in diesem Fall nach der Intensität der Ausprägung des numerischen Merkmals (Details finden sich in Kapitel 3.2). Dieses ist allgemein nicht zwingend üblich, oftmals werden die Säulen des Histogramms einfarbig abgebildet. Außerdem ermöglicht das Histogramm Aussagen über absolute Häufigkeiten in bestimmten Klassen zu tätigen. Darüber hinaus können beim Histogramm auch relative Häufigkeiten auf der Skala (y-Achse) aufgetragen werden, wie man im folgenden Beispiel (Abb. 5) sehen kann.

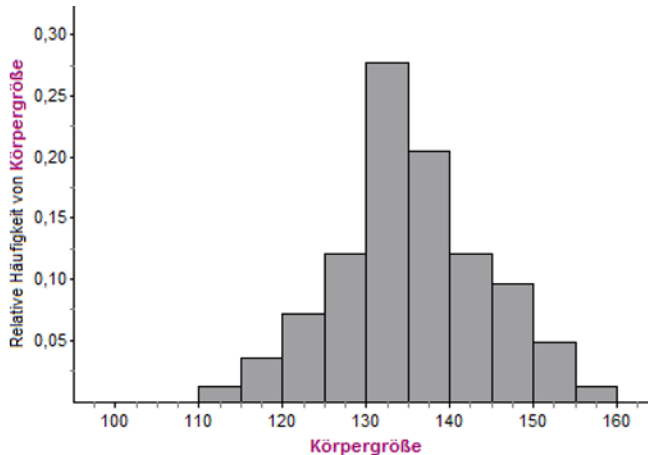


Abbildung 5: Verteilung des Merkmals „Körpergröße“ (Histogramm) (Datensatz Dettmar 2013)

Eine weitere prominente Möglichkeit ist die Darstellung der Verteilung des Merkmals Körpergröße als Boxplot (Abb. 6). Diese Darstellungsform gewährt Einblick in wichtige Charakteristika der Verteilungen.

Der Boxplot¹⁷ (Abbildung 6) reduziert die Verteilung auf eine Fünf-Kennzahlen-Zusammenfassung mit den Kennzahlen: Minimum, erstes Quartil, Median, drittes Quartil und Maximum. An dieser Darstellung lassen sich Aussagen zur Form (Lage der Box, Lage des Medians in der Box), zur Streuung (Breite der Box, Breite der Antennen), sowie zum Zentrum (Median) einer Verteilung machen und sie ist insbesondere dann hilfreich, wenn man zwei oder mehrere Verteilungen vergleichen möchte.

Wie bereits aus 2.1 bekannt ist, ist die Idee der Verteilung eine „Big idea“ und fundamentales Konzept nach Moore (1990). Rossman et al. (2001) unterscheiden sechs wichtige Aspekte zur Analyse von Verteilungen anhand von Leitfragen – diese finden sich in

¹⁷ Nähere Informationen zum Boxplot finden sich in Kütting (1994, 104-105).

Fahrmeir et al. (2007, 53) in nahezu identischer Weise wieder. Die fachlichen Aspekte und die Konzepte hinter den Verteilungen wie zum *Zentrum* (Fahrmeir et al. 2007, 53), zu den *Quantilen* (Fahrmeir et al. 2007, 64), sowie zur *Streuung* und *Schiefte* (Fahrmeir et al. 2007, 47f.) sollen an dieser Stelle nicht ausgeführt werden. Diese können bei den entsprechenden Literaturangaben nachgelesen werden.

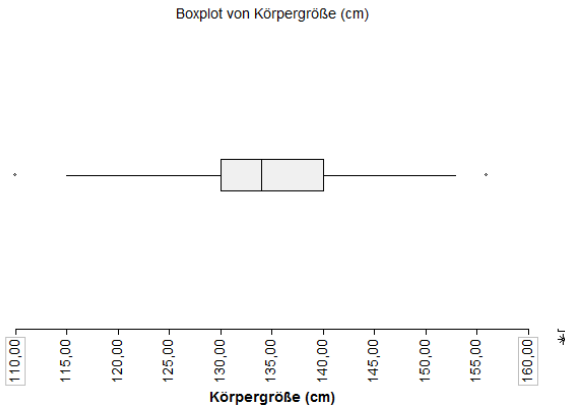


Abbildung 6: Verteilung des Merkmals „Körpergröße“ (Boxplot) (Datensatz Dettmar 2013)

Sichtweisen von Lernenden auf Verteilungen numerischer Merkmale

Es gibt mehrere Sichtweisen, die bei Lernenden charakterisiert werden können, wenn diese auf eine Verteilung eines numerischen Merkmals schauen und Eigenschaften der Verteilung beschreiben oder Fragestellungen zu dieser beantworten sollen. Es lassen sich innerhalb der Forschungsliteratur mehrere Konzeptualisierungen finden, die eine „lokale“ von einer „globalen“ Sichtweise oder ähnlich unterscheiden:

- Lokale Sicht („data as individual points“) vs. globale Sicht („data as entity“): (Bakker und Gravemeijer 2004)
- Lokale Sicht („local view“) vs. globale Sicht („aggregate view“) mit der Zwischenstufe der „Mini-aggregates“: (Makar und Confrey 2002 sowie Makar und Confrey 2005)
- „Data as pointer“, „Data as a focus on individual cases“, „Data as classifier“, „Data as aggregate“: (Konold et al. 2014)

Bakker und Gravemeijer (2004) unterscheiden (siehe Tabelle 1) zwischen der Sicht auf eine Verteilung als Ganzes und der Sicht auf einzelne Datenpunkte. Bakker und Gravemeijer (2004) listen in dieser Tabelle mögliche Aspekte auf, die Lernende nutzen können, um Verteilungen zu beschreiben. In ihrer Studie wurden Schüler der siebten Klasse untersucht, wie sie Verteilungen in Form von gestapelten Punktdiagrammen beschrei-

ben und auf welche Aspekte der Verteilungen sie verweisen. Bakker und Gravemeijer konstatieren nach ihren Beobachtungen, dass Lernende zum einen von der Sicht auf einzelne (Daten-)Punkte auf die Verteilung als Ganzes (mit ihren Eigenschaften wie Zentrum, Streuung, Dichte oder Schiefe) schauen und zum anderen auch auf ganzheitliche Eigenschaften (wie Zentrum, etc.) und von dort aus auf einzelne Fälle (individuelle (Daten-)punkte) schauen.

Tabelle 1: Verbindung zwischen Daten und Verteilung (Tabelle entnommen aus Bakker und Gravemeijer 2004)

Distribution (Conceptual Entity)				
Position and Shape (Global Informal Aspects)				
Center	Spread	Density	Skewness	
Mean, median, mode, midrange,...	Range, standard deviation, interquartile range,...	(relative) frequency, majority, outliers, ...	Position of data,...	majority
Data (plurality, individual data points)				

Generell stellen Bakker und Gravemeijer eine Entwicklung in der Perspektive der Lernenden von einzelnen Datenpunkten (Data, individual points) über modale Haufen („modal clumps“) bis zu einer Einteilung der Verteilung in die drei Bereiche „low“, „middle“ und „high“ fest.

Makar und Confrey (2005) machten die Beobachtung, dass es beim Blick auf die Daten mehr als nur zwei Sichtweisen (gesamtheitliche Sicht („aggregate view“) vs. punktweise Sicht („local view“)) gibt, nämlich als Zwischenstufe die Sichtweise „mini-aggregate“.

„...there are more than just the two perspectives of distribution that are usually discussed in the literature: single points and aggregate. This third perspective -partial distributions or “mini-aggregates”- deserves further research to investigate the strength of its link to statistical thinking about distributions.” (Makar und Confrey 2005, 48)

Makar und Confrey (2005) sprechen von einer „mini-aggregate“ Sichtweise, wenn ein bestimmtes Intervall, wie zum Beispiel eine Anhäufung von Daten in einer Verteilung, in den Blick genommen wird. Dies kann zum Beispiel ein „modaler Klumpen“ (engl. modal clump) sein. Auch Konold (2002b) nutzt die Beschreibung „modal clumps“, um eine Anhäufung von Daten in einer Verteilung zu beschreiben. Bakker (2004) fand in seiner Dissertation heraus, dass Lernende – anstatt sich nur auf die zentrale Region einer Verteilung zu konzentrieren – die Tendenz haben, die Daten in drei Teile einzuteilen („low“, „middle“ & „high“). „Modale Klumpen“ sind eine Möglichkeit, eine Vorstufe dieser Dreiteilung einer Verteilung.

Es werden insgesamt drei Sichtweisen auf Verteilungen charakterisiert: Zum einen wird von Bakker und Gravemeijer (2004) eine lokale Sichtweise von einer globalen Sichtweise auf Verteilungen unterschieden, zum anderen ergänzen Makar und Confrey (2005) diese Unterscheidung um die weitere Sichtweise, die der „Mini-Aggregates“.

Eine andere Beschreibung von Sichtweisen auf Verteilungen findet man bei Konold, Higgins, Russell und Khalil (2014). Konold et al. (2014) beschreiben in ihrem Artikel vier verschiedene Sichtweisen von Lernenden auf Verteilungen von Daten: „data as pointer“, „data as case value“, „data as classifier“ und „data as aggregate“. Sie führen das an einem Beispiel aus, indem sie eine Verteilung der Farben von sechs Kugeln (zwei grünen Kugeln, drei roten Kugeln und einer blauen Kugel) betrachten und verschiedene Aussagen von Lernenden zur Verteilung der Farben der sechs Kugeln kategorisieren (siehe Abbildung 7).

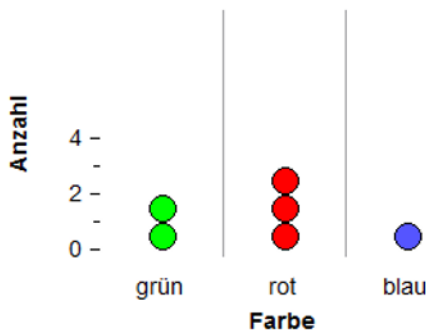


Abbildung 7: Verteilung von sechs farbigen Kugeln - entnommen aus Konold et al. (2014, 308) (eigene Darstellung)

Eine Aussage wie „We said our favourite colors“ („Wir haben unsere Lieblingsfarben gesagt“) zeigt keinen Bezug zur Verteilung des Merkmals Farbe bei den Kugeln und würde als Beispiel für „data as pointer“ gelten. „Juan liebt rot“ würde den einzelnen Fall (Juan) betonen und daher als „data as case value“ kategorisiert werden. Eine Aussage wie „drei Kinder lieben Rot“ würde man „classifier view on data“ und „die Hälfte der Kinder liebt rot“ „aggregate view on data“ nennen. Den Unterschied zwischen „aggregate view“ und „classifier view“ macht Konold et al. (2014) an einem Beispiel klar: Ein Statistiker würde auf die Daten mit einem gesamten Blick („aggregate“) draufschauen. Im Gegensatz dazu würde man bei der „classifier“-Perspektive nur Daten mit dem gleichen Wert betrachten. Wird ein Fall nur einzeln in den Blick genommen, so spricht man hier von der „case value“-Perspektive. Diese hier beschriebenen Perspektiven werden von den Autoren als „bewegliche“ Hierarchie betrachtet, bei denen eine „höhere Stufe“ die niedrigeren fasst“. Konold et al. (2014) betonen, dass das Zusammenfassen in-

dividueller Daten durch einen Wert oder durch eine gesamte Verteilung Lernenden nicht leicht fällt.

Eine besondere Form, die Verteilung numerischer Merkmale darzustellen, bieten Boxplots. Diese Darstellungsmöglichkeit von Verteilungen numerischer Merkmale wollen wir im Folgenden kurz beschreiben.

Boxplots als Diagramme – Eine besondere Form, Verteilungen numerischer Merkmale darzustellen

Die Anfänge der Verwendung von Boxplots in der deskriptiven Statistik finden sich bereits 1970 in einer Arbeit des US-Statistikers Tukey. Dieser Boxplot wurde aber erst sieben Jahre später, 1977, vorgestellt und veröffentlicht. In diesem sogenannten Tukey-Boxplot finden sich neben Lagemaßen auch Streuungsmaße zu Verteilungen eines quantitativen Merkmals wieder. Eine Übersicht über verschiedene Variationen von Boxplots findet sich in Wickham und Stryjewski (2011, 5-9). Wir wollen uns im Folgenden mit dem Boxplot nach Tukey auseinandersetzen. Verteilungen numerischer Merkmale lassen sich durch die fünf Tukey'schen Zahlen (Biehler 1982, 42-54) beschreiben: Minimum, 1. Quartil, Median, 3. Quartil und Maximum. Diese können im sogenannten Boxplot abgebildet werden. Es gibt mehrere Definitionsmöglichkeiten für Boxplots (vor allem, wenn man die Definition möglicher Ausreißer¹⁸ in einem Datensatz mit in den Blick nimmt). Deshalb werden im Folgenden und auch im weiteren Verlauf dieser Arbeit nur Boxplots nach der Tukey'schen Definition und Darstellung betrachtet. Die Abbildung 8 zeigt die Verteilung eines numerischen Merkmals (Bruttomonatsverdienst) aus dem Datensatz der Verdienststrukturerhebung 2006¹⁹ dargestellt als Boxplot.

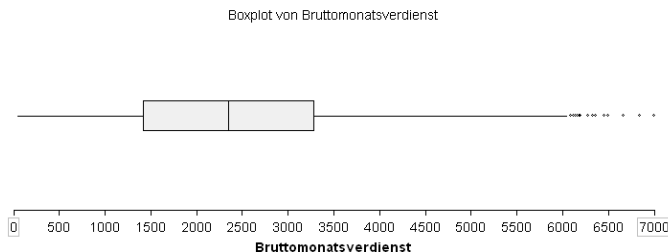


Abbildung 8: Tukey-Boxplot in TinkerPlots zur Verteilung des Merkmals Bruttomonatsverdienst (Datensatz: Verdienststrukturerhebung 2006)

18 Die Grenzen für die Ausreißer sind dabei so definiert: $fu = Q1 - 1,5 \cdot (Q3 - Q1)$ sowie $fo = Q3 + 1,5 \cdot (Q3 - Q1)$. Dabei werden die Antennen bis zu den Datenpunkten, die gerade noch größer als fu bzw. gerade noch kleiner als fo gezeichnet. Werte, die außerhalb dieser Bereiche liegen, werden als Punkte gekennzeichnet.

19 Dieser Datensatz wird ausführlich in Kapitel 9.2.1 beschrieben.

Der Boxplot bietet eine Zusammenfassung dieser fünf, oben erwähnten Werte und ermöglicht es so, Verteilungen auf diese fünf Werte beschränkt zu beschreiben und zu charakterisieren. Biehler (2007c, 2-3) hat verschiedene Gebrauchsweisen von Boxplots charakterisiert.

- *Boxplot als Zusammenfassung* (location summary): In diesem Fall wird der Boxplot als Zusammenfassung der fünf Tukey'schen Kennzahlen angesehen. Vorsicht ist geboten bei den Quartilen, weil es diesbezüglich verschiedene Definitionen gibt (Langford 2006). Als Sprechweisen nimmt man in diesem Fall „Boxplot als Zusammenfassung“: „Ungefähr 50% der Daten sind in der Box, bzw. 50% der Daten liegen zwischen Q1 und Q3.“ [...]
- *Boxplot, um Daten zu klassifizieren*: Der Boxplot „teilt“ die Daten in vier Intervalle. Oft, so Biehler (2007c) und Bakker (2004), sehen Lernende eine Einteilung in drei Intervalle: unteres Viertel, Mitte (50%) und oberes Viertel, wobei die „Mehrheit“ der Daten dann in der Mitte liegt.
- *Regionale durchschnittliche Dichte in den vier Intervallen*: Bei dieser Gebrauchsweise wird vor allem auf die Dichte der Daten eingegangen, insbesondere, dass beim Boxplot die Dichte der Daten antiproportional zu den Längen der vier Intervalle ist. Biehler (2007c) spricht in diesem Fall von „regionaler“ Dichte, weil sie weder global noch lokal zu betrachten ist.
- *Regionale Streuung*: In diesem Fall wird vorgeschlagen, regionale Maße für Streuung zu erschaffen: z.B. Differenz aus Minimum und Median, Differenz aus Median und Maximum. Als Sprechweise für die Lernenden wird vorgeschlagen von der „Streuung der mittleren Hälfte der Daten“ und von der „Streuung des ersten Viertels“ zu sprechen. Der bekannte Interquartilsabstand kann als globales Lagemaß für Streuung angesehen werden, ähnlich wie die Standardabweichung.
- *Boxplots, die Streuung links und rechts vom Median zeigen*: In diesem Fall erfolgt meist eine Umdeutung des 1. und 3. Quartils: Das erste Quartil wird als Median von Minimum und Median, das dritte Quartil in diesem Falle als Median von Median und Maximum des Datensatzes gedeutet. Welche Schlüsse lassen sich daraus ziehen? Die Differenz beispielsweise aus dem dritten Quartil und dem Median ist die Abweichung vom Median nach oben, so sind zum Beispiel ungefähr die Hälfte der Abweichungen von den höheren Werten größer als die Differenz Q3-Median und ungefähr die Hälfte der Abweichungen sind kleiner als Q3-Median. Man kann somit zwei verschiedene globale Lagemaße für Streuung dem Boxplot entnehmen: Median-Q1 als Abweichung kleinerer Werte vom Median und Q3-Median als Abweichung größerer Werte vom Median. [...] Genauso kann man den Bereich / die Zahl auch als regionales Streuungsmaß deuten.
- *Boxplots als „Mitte und Streuung“ Indikator*: In diesem Fall werden Boxplots genutzt, um sowohl mit der Mitte (in diesem Fall der Median) als auch mit der Streuung (in diesem Fall der Interquartilsabstand) zu argumentieren. Die Asymmetrie der Box kann die Gestalt der Verteilung suggerieren.²⁰ (übersetzt aus Biehler 2007c, 2-3)

Es bleibt festzustellen, dass Boxplots sehr abstrakt und konzeptreich sind und von Schülern wie auch von Lehrern oft nur sehr schwer zu verstehen sind (vgl. Bakker, Biehler und Konold 2005). Bekannte Fehlvorstellungen und Konzepte sind nach Bakker et al. (2005), dass der Boxplot den Datensatz in genau vier gleich große Bereiche teilt, genau

20 Anmerkung von Biehler (2007c): Der Begriff der regionalen Streuung, sowie die „untere“ und „obere“ Abweichung sind in der Statistik unüblich, als Zwischenschritt, so Biehler (2007c), können sie aber dennoch nützlich sein, wenn z.B. ein Verständnis von Besonderheiten von Verteilungen entwickelt werden soll (vgl. z.B. Konold et al. 2002).

50% der Daten in der Box liegen und desto breiter die Box ist, desto mehr Daten in ihr liegen.

Auch Lem et al. (2014) bestätigen die Beobachtungen von Bakker et al. (2005) hinsichtlich der Schwierigkeiten beim Interpretieren von Boxplots. Zusätzlich zu Bakker et al. (2005) konstatieren sie:

„Students have more difficulties interpreting box plots than other external representations of data distributions, such as histograms or descriptive statistics (Lem et al. 2013a). Common misinterpretations are, for example, thinking that the median line actually represents the mean, or ignoring the whiskers in the assumption that no data are represented in this part of the box plot (e.g. Lem et al. 2012, 2013a).“ (Lem et al. 2014, 1)

Wie oben schon erwähnt, bietet der Boxplot eine ganzheitliche Sicht auf Daten an, da er aus den fünf Tukey'schen Zahlen konstruiert wird. Dennoch sind Lernende verleitet, Einzelfälle abzuleiten. Ebenso stellen die im Boxplot ablesbaren Dichteunterschiede in den Daten eine Hürde für Lernende dar. Die Situation, dass „bei einer schmalen Box, die Daten in diesem Intervall umso dichter liegen“ stellt sich für viele Lernende als kognitive Herausforderung dar. (Bakker et al., 2005) Ähnliches stellt auch Biehler (1997a, 37-38) fest:

„Gravierender sind die Probleme im Boxplot, wo abweichende Graphikkonventionen angewendet werden, die offensichtlich nicht genügend im Unterricht reflektiert wurden. Beispielsweise bedeutet eine größere Fläche der Box im Boxplot eine geringere Datendichte und nicht eine größere Häufigkeit wie im Histogramm. Es gibt Hinweise darauf, daß eine zu schnelle Übernahme fertiger in Software angebotener Graphiken, ohne über deren Konstruktion zu reflektieren, hierfür verantwortlich sein könnte. Jenseits dieser schwierigen Graphikkonventionen zeigten sich begriffliche und sprachliche Schwierigkeiten bei der Formulierung von Beobachtungen, die die Lernenden in den Graphiken gemacht hatten. Ein besonders schwieriges Problem betrifft die Häufigkeitsverteilung (Intervall -> Häufigkeit) und ihre "Umkehrung", die Zuordnung von kumulierten Häufigkeiten zu Intervallen (Quantilbegriff), wie sie exemplarisch im Boxplot vorgenommen wird.“ (Biehler 1997a, 37-38)

Insgesamt gibt es aber (trotz der konzeptionellen Schwierigkeiten) Vorteile (wie die Reduzierung der Daten auf die fünf Kennzahlen nach Tukey) den Boxplots zu nutzen, um Verteilungen numerischer Merkmale darzustellen. Diese Vorteile werden besonders beim Vergleich von Verteilungen eines numerischen Merkmals deutlich.

2.3.2 Verteilungsvergleiche

Im folgenden Abschnitt wird beschrieben, was in dieser Arbeit unter einem Verteilungsvergleich zu verstehen ist und welche verschiedenen Formen dieser annehmen kann. Im Weiteren sollen dann verschiedene Möglichkeiten aufgezeigt werden, Verteilungen zu vergleichen. Dabei orientiert sich diese Arbeit auf das Herausarbeiten von Unterschieden zwischen Verteilungen anhand der von Rossman et al. (2001) definierten Charakteristika von Verteilungen (siehe 2.1.2), die als mögliche Verteilungsvergleich-

Elemente angesehen werden. Darüber hinaus werden weitere mögliche Strategien von Lernenden beim Vergleichen von Verteilungen aufgeführt. Mit Gruppenvergleiche (beziehungsweise Verteilungsvergleiche) sind in dieser Arbeit Untersuchungen gemeint, die von Fragestellungen wie „Inwiefern unterscheiden sich die befragten Mädchen von den befragten Jungen hinsichtlich ihres Fernsehkonsums?“ oder „Lesen die befragten Mädchen mehr als die befragten Jungen (Stunden pro Woche)?“ eingeleitet werden.

Man sieht an diesen Beispielen, dass es unterschiedliche Qualitäten von Fragestellungen gibt, die zu einem Gruppenvergleich führen.²¹ Während die erste Frage dazu anregt, Unterschiede in den beiden Gruppen herauszuarbeiten (z.B. hinsichtlich der Mittelwerte, der Streuung, der Verteilungsform, usw.), verleitet die zweite Frage zu einer Ja/Nein-Antwort, vielleicht anhand des Unterschiedes der arithmetischen Mittelwerte des Merkmals Zeit_Lesen in den beiden Gruppen. Man kann Gruppenvergleiche aber auch anhand von bekannten Beispielen aus der medizinischen oder auch aus der empirischen Forschung motivieren:

„Most of the important issues and questions argued with data amount to comparing two groups, for example, treatment and control groups in medicine. Before-and-after groups in various interventions and educational studies, and females versus males in gender equity studies.“ (Konold und Higgins 2003, 207)

Oder man stellt sich Verteilungsvergleiche in verschiedenen („Anwendungs“-) Kontexten vor, wie zum Beispiel: „X is larger in group A than in group B“ oder auch „Group comparison in a decision context“ nach dem Motto: Welche Gruppe ist „besser“? (siehe auch Biehler 2007c). So lassen sich zusammenfassend die folgenden Typen von Fragestellungen finden, die einem Verteilungsvergleich vorangehen können:

- Typ1: Entscheidungsfragestellungen: Welche Gruppe ist besser?
- Typ2: Explorative Fragestellungen: Welche Unterschiede/Gemeinsamkeiten können identifiziert werden?
- Typ3: Hypothesen-gestützte Fragestellungen: Ist das arithmetische Mittel in Gruppe 1 größer als in Gruppe 2? Tendieren Jungen dazu mehr Zeit pro Woche in Stunden am PC zu verbringen als Mädchen?

Konold et al. (1997, 7) unterscheiden im Weiteren verschiedene Szenarien hinsichtlich der Variablen („kategorial & kategorial“, „kategorial & numerisch“ sowie „numerisch & numerisch“) beim Vergleich zweier Merkmale:

- „Scenario1: Comparison involving two categorical (cat) variables (“Are males or females more likely to have a driver’s license?”)
- „Scenario2: Comparison involving one numeric (num) and one categorical variable (“Do those with a curfew tend to study more hours than those without a curfew?”)

21 Für eine genauere Unterscheidung siehe Biehler (2001, 98).

- Scenario3: Comparison involving two numeric variables (“Is there a relation between hours spent watching TV and school grades?”) (vgl. Konold et al. 1997, 7)

Im Folgenden befassen wir uns ausschließlich mit Vergleichen vom Szenario 2.

Wie Rossman et al. (2001) aufgreifen, lassen sich Verteilungen anhand ihres Zentrums, ihrer Streuung, ihrer Form, anhand von Teilgruppen, Ausreißern und anhand ihrer Struktur charakterisieren. Elemente von Rossman et al. (2001) finden sich auch bei Zieffler, Harring und Long (2011) wieder. Zieffler et al. (2011) unterscheiden bei Vergleichsgrößen zwischen „measures of location, or central tendency“ und „measures of variability, or dispersion“ (vgl. Zieffler et al. 2011, 74) und zeigen darüber hinaus auch Vergleichsmöglichkeiten anhand der Formen der Verteilungen („Skewness“) auf (vgl. Zieffler et al. 2011, 78).

Diese Charakteristika werden als Ausgangspunkt für mögliche Vergleichsaspekte genommen, welche im Folgenden durch weitere Ansätze (vgl. Biehler 2001 und Biehler 2007b) angereichert werden. So könnte man auf einer ersten Ebene identifizieren, welcher Mittelwert (arithmetisches Mittel / Median) verwendet wurde, und ob weitere oder andere Mittelwerte andere Aussagen ermöglichen (vgl. Biehler 2007b, 4). Über die Mittelwerte hinaus sollte man entscheiden, ob es auch noch weitere Unterschiede zwischen den Verteilungen gibt und ob sich die Verteilungen vielleicht auch in Streuung und Form unterscheiden. Auf einer weiteren Ebene könnte man sich dann, so Biehler (2007b), fragen, ob man die Unterschiede noch anders herausarbeiten kann, wenn man jeweils die Gruppe der „Vielspieler“²² vergleicht („mehr als 10 Stunden pro Woche; wie viel Prozent der Mädchen, wie viel Prozent der Jungen spielen mehr als 10 Stunden?). Hier kann man in Anlehnung an Biehler (2001, 110) zwischen h- und q- basierten Vergleichen unterscheiden. Während h-basierte Vergleiche der Fragestellung nachgehen „Wie viel Prozent der Jungen spielen mehr als 10 Stunden im Vergleich zur Gruppe der Mädchen?“ ist es bei q-basierten Vergleichen genau umgekehrt. Hier würde eine Frage etwa so lauten: „Wie lang ist die Zeit, die am Computer verbracht wird, bei den oberen 10% mindestens?“. Dieses wäre ein Vergleich anhand des 90%-Quantils.

Die konkrete Definition eines h-basierten Vergleichs²³ lautet bei Biehler (2001, 110) wie folgt:

„Vergleiche zwischen zwei statistischen Variablen nenne ich h-basiert, 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, 110)

-
- 22 Der Begriff „Vielspieler“ bezieht sich auf Untersuchungen zum geschlechtsspezifischen Computernutzungsverhalten im Muffins-Datensatz (siehe Biehler et al. 2003). In diesem Fall sind Schüler gemeint, die 10 oder mehr Stunden pro Woche am Computer spielen.
- 23 Anstelle von h(äufigkeits)-basierten Vergleichen sprechen wir im Folgenden von p(roportional)-basierten Vergleichen.

Einen q -basierten Vergleich beschreibt Biehler (2001, 110) so:

„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, 110)

Weiterhin kann es nützlich sein (siehe z.B. Konold et al. 2002), so genannte modale Haufen („modal clumps“) in Verteilungen zu identifizieren und daran anschließend eine Verschiebung zwischen zwei Verteilungen identifizieren. Auf einer elaborierteren Ebene kann dann die Ermittlung der Verschiebung zwischen den beiden Verteilungen (anhand der Unterschiede der jeweiligen fünf Tukey-Kennzahlen) mit Hilfe eines Shift-Modells erfolgen (vgl. Biehler 2007c). Hier kann man zwischen einem additiven Shift-Modell und einem multiplikativen Shift-Modell unterscheiden. An diesem lässt sich darüber hinaus klar machen, dass die Verschiebung zwischen zwei Verteilungen sich nicht nur auf das arithmetische Mittel sondern auch auf die Verteilung als Ganzes beziehen kann, aber nicht muss. Unterschiede bezüglich der Streuung zwischen beiden Verteilungen können anhand von Boxplots herausgearbeitet und präzisiert werden. Vergleicht man beispielsweise die Interquartilsabstände zweier Verteilungen so kann ein größerer Interquartilsabstand in der einen Verteilung ein heterogeneres Verhalten der entsprechenden Gruppe im Vergleich zu der anderen Verteilung andeuten. Unterschiedliche Darstellungen bieten (auch im Sinne einer explorierenden Haltung) unterschiedliche Einsichten in die vorliegenden Daten und Verteilungen. Daher kann es zweckmäßig sein, zwischen mehreren Verteilungen zu switchen, um Muster in den Daten zu entdecken.

Welche tragfähigen Möglichkeiten kann man nun festhalten, um Verteilungen zu vergleichen? Nimmt man die Elemente von Rossman et al. (2001) sowie die Analysen von Biehler (2007b) und Biehler (2007c), so kann man die folgenden tragfähigen Vergleichskonzepte (im Folgenden „Verteilungsvergleich-Elemente“) beim Vergleich von Verteilungen eines numerischen Merkmals in der deskriptiven Statistik identifizieren (siehe Tabelle 2).²⁴ Um einen Einblick in die Anwendung der in der Tabelle 2 aufgeführten Verteilungsvergleich-Elemente zu bekommen, nutzen wir dazu den Datensatz von Dettmar (2013), der Daten von 91 Kindern einer Grundschule enthält²⁵ und gehen der Fragestellung „Inwiefern unterscheiden sich die Schülerinnen und Schüler hinsicht-

24 Dabei soll die Konzentration auf Verteilungsvergleich-Elemente als solche, nicht aber auf mögliche Qualitätsunterschiede bezüglich des Herausarbeitens von Unterschieden eingegangen werden. Auf Qualitätsunterschiede beim Herausarbeiten von Unterschieden gehen wir ausführlich in Kapitel 6 und in Kapitel 10 ein.

25 Die Diagramme wurden in diesem Abschnitt mit der Software TinkerPlots erstellt. Auch die jeweiligen Kennzahlen wurden mit Hilfe von TinkerPlots berechnet. Auf einzelne Details zum Erstellen dieser hier abgebildeten Graphiken sowie auf das Vergleichen von Verteilungen mit TinkerPlots werden wir in Kapitel 3 ausführlich eingehen.

lich ihrer Körpergröße?“ nach. Dabei wird versucht, anhand der Verteilungsvergleich-Elemente (aus Tabelle 2), Unterschiede zwischen den Verteilungen herauszuarbeiten.²⁶

Tabelle 2: Übersicht über die einzelnen Elemente zum Vergleich von Verteilungen

Verteilungsvergleich-Element	Beschreibung
Zentrum_aMittel	Es werden zwei Verteilungen eines numerischen Merkmals anhand der Unterschiede zwischen ihren arithmetischen Mittelwertes verglichen.
Zentrum_Median	Es werden zwei Verteilungen eines numerischen Merkmals anhand der Unterschiede zwischen ihren Medianen verglichen. Auf einer informellen Ebene würde man hier auch den Vergleich der Lage zweier modaler Klumpen („modal clumps“) als Vergleich der Zentren zweier Verteilungen auffassen.
Streuung	Es werden zwei Verteilungen eines numerischen Merkmals anhand der Unterschiede zwischen ihren Streumaßen verglichen. Mögliche Streumaße sind: <ul style="list-style-type: none"> • Spannweite • IQR (Interquartilsabstand) • Informelle Beschreibungen der Streuung (wie „dichter“, etc.) • Standardabweichung des arithmetischen Mittels
Verschiebung	Es werden zwei Verteilungen eines numerischen Merkmals anhand der Charakterisierung einer Verschiebung zwischen beiden Verteilungen verglichen. ²⁷
Form	Es werden zwei Verteilungen eines numerischen Merkmals anhand der Unterschiede zwischen der Form beider Verteilungen verglichen. (z.B. Identifizierung von Unterschieden bei der Schiefe oder Symmetrie der Verteilungen)
p-basiert	Es werden zwei Verteilungen eines numerischen Merkmals p-basiert verglichen.
q-basiert	Es werden zwei Verteilungen eines numerischen Merkmals q-basiert verglichen.

Vergleich von Verteilungen anhand von Mittelwerten (arithmetisches Mittel / Median)

Eine erste Möglichkeit ist gegeben, indem man die Mittelwerte der Verteilungen des Merkmals Körpergröße zwischen Schülerinnen und Schülern in dem Datensatz von Dettmar (2013) vergleicht (siehe Abb. 9). Hier liegt kein Unterschied zwischen den arithmetischen Mittelwerten vor, d.h., dass sowohl Jungen als auch Mädchen in diesem Datensatz im Durchschnitt gleich groß sind. Für den Vergleichsaspekt „Zentrum“ wird daher ausnahmsweise der Muffins-Datensatz (Biehler, Kombrink und Schweynoch

26 Dabei bleibt zu bemerken, dass es sich jeweils um Vergleichsaussagen innerhalb der gegebenen Stichprobe handelt. Inferenzielle Schlüsse über die Stichprobe hinaus sollen hier nicht vorgenommen werden.

27 Für eine detailliertere Ausführung siehe unten (in diesem Abschnitt).

2003), der die Angaben zum Freizeitverhalten und zum Medienkonsum von 538 Schülern aus elften Jahrgangsstufen in NRW enthält, genommen.

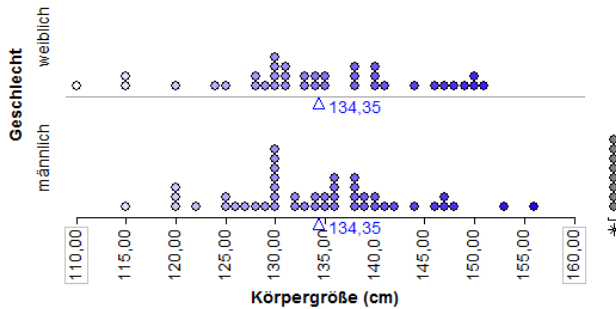


Abbildung 9: Beispiele zum Verteilungsvergleich „Vergleich anhand des arithmetischen Mittelwertes“ (Datensatz aus Dettmar 2013)

In diesem können deutliche Unterschiede bezüglich der Mittelwerte hinsichtlich von Schülerinnen und Schülern identifiziert werden (siehe Abb. 10).

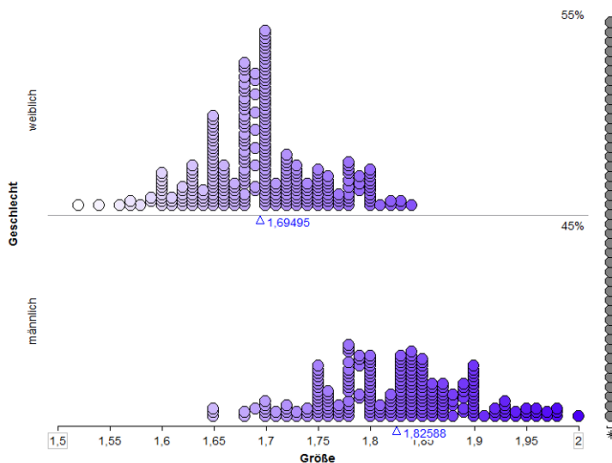


Abbildung 10: Beispiele zum Verteilungsvergleich „Vergleich anhand des arithmetischen Mittelwertes“ (Datensatz: Muffins)

Werden nur die beiden arithmetischen Mittelwerte (der Verteilungen der Jungen und Mädchen) angegeben, ist das noch kein konkreter Vergleich. Vielmehr lässt sich anhand der arithmetischen Mittelwerte beider Verteilungen in diesem Beispiel sagen, dass die

Schüler im Muffins-Datensatz durchschnittlich ca. 13 cm größer sind, als die Schülerinnen. Dieses könnte man als „additive“ Vergleichsaussage bezüglich des Unterschieds der arithmetischen Mittelwerte bezeichnen. Ebenfalls lässt sich diese Aussage auch „multiplikativ“ formulieren: Die Schüler im Muffins-Datensatz sind im Durchschnitt um ca. 8% größer als die Schülerinnen. Beim arithmetischen Mittel lassen sich noch weitere verschiedene Aspekte beim Vergleich der arithmetischen Mittelwerte beider Verteilungen geltend machen. (für verschiedene Interpretationsweisen des arithmetischen Mittel, z.B. beim Verteilungsvergleich, siehe Konold und Pollatsek 2002, 270f.). Bei Datensätzen, die gleich groß sind, können Verteilungen anhand des „Total score“ (Gesamtsumme der Daten in der jeweiligen Verteilung) verglichen werden, wenn die Anzahl der Fälle in beiden Gruppen gleich ist. Dieses bietet sich an, wenn beispielsweise Schulklassen anhand ihrer in einem Test erreichten Punkte verglichen werden. Man kann so in beiden Verteilungen die Punkte aufaddieren und dann die Gesamtpunktzahl vergleichen. Dieses Vorgehen funktioniert allerdings nur bei gleich großen Gruppen. Sind die Klassen von der Anzahl her nicht gleich groß, so muss das arithmetische Mittel als Vergleichsmaßstab genutzt werden.

Beim Median (Abb. 11) lassen sich ebenfalls wie beim arithmetischen Mittel Vergleichsaussagen auf einer „additiven“ sowie auf einer „multiplikativen“ Weise tätigen.

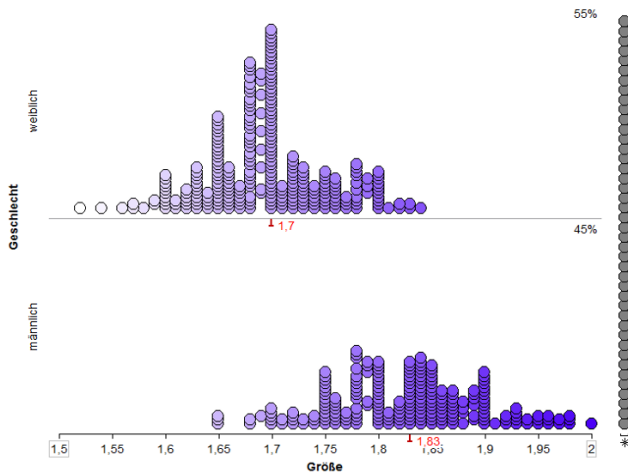


Abbildung 11: Beispiele zum Verteilungsvergleich „Vergleich anhand des Median“ (Datensatz Muffins)

So kann anhand dieser Verteilungen festgehalten werden, dass die Schüler des Muffins-Datensatzes im Median 13cm (bzw., ca. 7,1%) größer sind als die Schülerinnen.

An dieser Stelle kann es sich anbieten, Unterschiede zwischen den arithmetischen Mittelwert-Unterschieden und den Median-Unterschieden in Beziehung zu setzen. Hier soll allerdings nicht explizit darauf eingegangen werden (da es uns in diesem Abschnitt primär darum geht, tragfähige Konzepte beim Vergleichen von Verteilungen herauszuarbeiten).

Vergleich von Verteilungen anhand der Streuung

Vergleicht man die Verteilungen anhand der Streuung, so gibt es zunächst zwei formale Möglichkeiten dieses durchzuführen. Zum einen anhand der Spannweite, zum anderen anhand der mittleren 50% (Interquartilsabstand). In dem uns hier vorliegenden Beispiel (Abb. 12), die Verteilungen des Merkmals Körpergröße unterschieden nach dem Merkmal Geschlecht aus dem Datensatz von Dettmar (2013), sind die Spannweiten beider Verteilungen gleich (41,0 cm), die Breite der Boxen unterscheidet sich jedoch ein wenig. Während der Interquartilsabstand der Verteilung des Merkmals Körpergröße bei den Schülerinnen 10 cm beträgt, beträgt er bei den Schülern 9 cm. Man könnte bei den Schülerinnen von einem im Vergleich zu den Schülern leicht heterogenerem Verhalten sprechen.

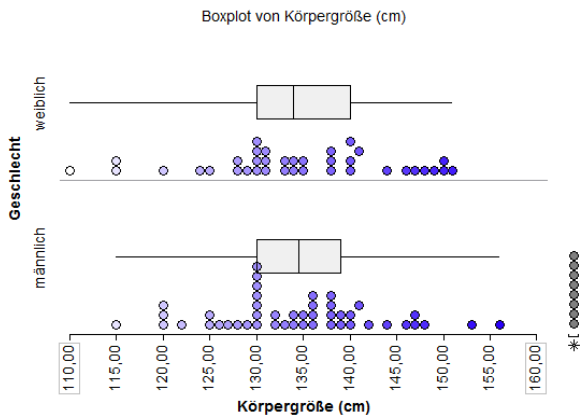


Abbildung 12: Beispiele zum Verteilungsvergleich „Vergleich anhand der Streuung“ (Datensatz aus Dettmar 2013)

Vergleich von Verteilungen anhand der Form

Verteilungen lassen sich auch anhand der Form (oder Schiefe) vergleichen. Wenn eine Verteilung als rechtsschief (linkssteil) beschrieben würde und die andere als linksschief (rechtssteil) so hätte man gravierende Unterschiede bezüglich der Form beider Verteilungen herausgefunden und könnte somit auch Unterschiede zwischen den Gruppen feststellen.

Vergleich von Verteilungen anhand einer Verschiebung

Eine weitere elaborierte Vergleichsmöglichkeit beim Vergleich zweier Verteilungen liegt darin, zu prüfen, ob eine Verschiebung der gesamten Verteilung vorliegt (siehe z.B. Abb. 13).

Die Verschiebung zwischen zwei Verteilungen lässt sich vereinfacht mit der „5-Zahlen-Zusammenfassung“ feststellen. Eine Vergleichsaussage anhand dieser „5-Zahlen-Zusammenfassung“ ist einfach, wenn alle fünf Kennzahlen der einen Verteilung größer sind, als die der anderen Verteilung. Man kann überprüfen, ob die fünf Zahlen (Minimum, Q1, Median, Q3 und Maximum) näherungsweise um einen festen Betrag, bzw. um einen festen Faktor verschoben sind.

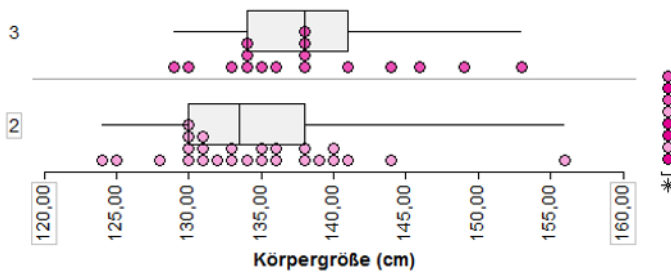


Abbildung 13: Beispiele zum Verteilungsvergleich „Vergleich durch Feststellung einer Verschiebung“ (Datensatz aus Dettmar 2013)

Man spricht dann von einer additiven bzw. multiplikativen Verschiebung der Verteilung. Mathematischer Hintergrund ist der Begriff „stochastisch größer“, wie er bezogen auf Zufallsgrößen verwendet wird. Biehler (2001) überträgt das in die beschreibende Statistik:

„ V heißt „stochastisch größer“ als W , wenn alle Quantile von V größer oder gleich als alle entsprechenden von W sind. Praktisch äquivalent dazu ist, dass für alle x die Beziehung $h(V \leq x) \leq h(W \leq x)$ gilt (Genau genommen gilt dies nur für ideale stetige Verteilungen, vgl. Pfanzagl 1991). Eine sinnvolle Mathematisierung der intuitiven Vorstellung, V ist größer als W bzw. V in Gruppe 1 ist größer als V in Gruppe 2 ist mit diesem Konzept gegeben, das man graphisch sehr gut an der kumulativen Verteilungsfunktion festmachen kann. Ist in diesem Sinne eine Variable stochastisch größer als die andere müssen alle 5 im Boxplot dargestellten Kennzahlen größer sein. Ist das nicht der Fall, so müssen differenzierende Aussagen getroffen werden“ (Biehler 2001, 110)

„Stochastisch Größer“ impliziert ein entsprechendes Verhalten der 5 Kennzahlen. Die Umkehrung gilt natürlich nicht, aber es erscheint legitim, auf die 5 Kennzahlen im Sinne einer modellhaft vereinfachenden Beschreibung der Beziehung zwischen zwei Verteilungen zurückzugreifen. Biehler (2007c) nennt dies das „Shift-Modell“. Ein Shift-

Modell bietet die Möglichkeit eine Verschiebung zwischen Verteilungen weiter auszu-differenzieren, indem man u.a. zwischen einer additiven Verschiebung und einer multiplikativen Verschiebung und einer gleichmäßigen (uniform) und ungleichmäßigen Verschiebung unterscheidet (siehe Biehler 2007c):

„If the additive shift model holds, we have X has the same distribution as $Y + a$, and if $a > 0$ this implies that X is stochastically larger than Y and that the properties hold $Q_p(X) \geq Q_p(Y) + a$ for all $p \in (0;1)$. If $a < 0$, Y is stochastically larger than X . Thus the shift model is a special case of being “stochastically larger”. Another special case is the multiplicative shift model: X has the same distribution as $a \cdot Y$, $a > 0$. If $a > 1$, X is stochastically larger than Y and all the quartiles multiply by the same factor a , which is also the factor by which any measure of spread increases.”

Biehler (2007c, 7) führt abschließend vier mögliche Typen der Verschiebung der Verteilungen auf.

- „ X is (statistically larger) in group A than in group B: we speak of a shift to higher values
- The distribution of X in group B can be described as an uniform additive shift of the distribution of X in group A
- The distribution of X in group B can be described as an multiplicative shift of the distribution of X in group A
- The difference in distribution between both groups is more complex” (Biehler 2007c, 7)

Vergleich von Verteilungen anhand p-basierter Vergleiche

Es lassen sich z.B. „große“ Schüler definieren und die Anzahlen der „großen“ (z.B. Schüler, die 145,00cm oder größer sind) Schüler in beiden Verteilungen bestimmen und vergleichen (siehe Abb. 14). In diesem Fall würde man bei den weiblichen Schülerinnen sieben „große“ und bei den männlichen sechs „große“ Kinder finden. Eine Vergleichbarkeit ist aufgrund der unterschiedlichen Anzahlen in beiden Verteilungen anhand absoluter Häufigkeiten aber nicht gegeben. Somit ist diese Strategie nur tragfähig, wenn in beiden Gruppen gleich viele Fälle vorhanden sind.

Nimmt man nun allerdings relative Häufigkeiten und bestimmt in beiden Verteilungen den Anteil der „größeren Kinder“ in der jeweiligen Verteilung, so vergleicht man die Anteile der „Großen“ in beiden Klassen, eine Vergleichsart, die Biehler (2001, 110) „p-basiert“ (bzw. „h-basiert“) nennt:

„Vergleiche zwischen zwei statistischen Variablen nenne ich h-basiert, 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, 110)

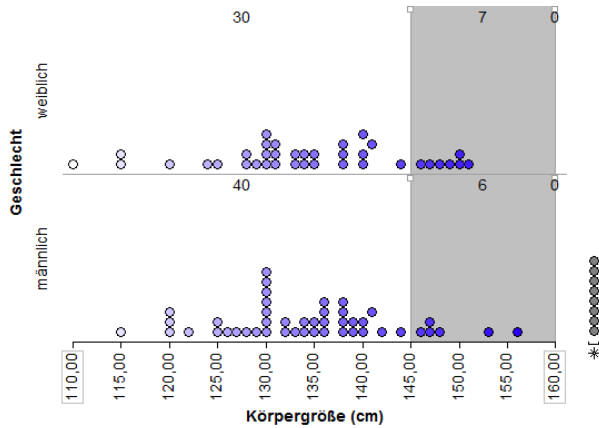


Abbildung 14: Beispiel zum Verteilungsvergleich: „Vergleichen gleich großer Klassen anhand von absoluten Häufigkeiten“ (Datensatz aus Dettmar 2013)

Wir bestimmen den Anteil (relative Häufigkeit) der Fälle in besagten Intervallen und stellen im vorliegenden Beispiel fest, dass in diesem Fall 19% der Schülerinnen und 13% der Schüler „groß“ sind (Abb. 15).

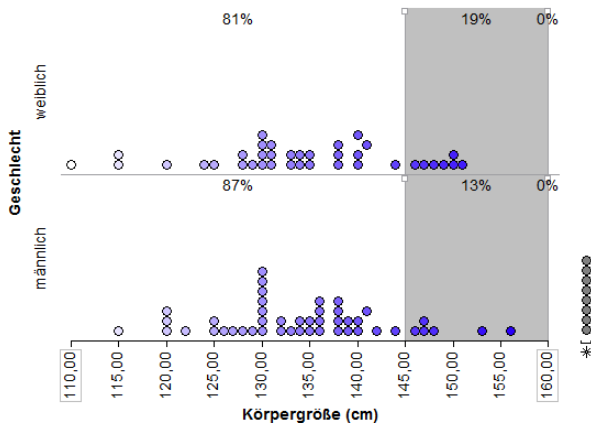


Abbildung 15: Beispiel zum Verteilungsvergleich „Vergleichen gleich großer Klassen anhand von relativen Häufigkeiten“ (p-basiert) (Datensatz aus Dettmar 2013)

Dieses würden wir als tragfähiges Vergleichskonzept betrachten. Es bleibt zu bemerken, dass die Definition der Körpergröße „groß“ willkürlich ist. Anstatt zu sagen, dass Schü-

ler, die größer oder gleich 1,45m sind, „groß“ sind, hätte man dieses Kriterium auch auf Schüler, die größer oder gleich 1,40m sind, übertragen können.

Es bleibt die Frage: Was ist, wenn einzelne Intervalle aus den Verteilungen herausgegriffen werden, die sich nicht in „große Schüler“ bzw. „kleine Schüler“ einteilen lassen? Zum Beispiel (siehe Abb. 16) wäre es ja vorstellbar, dass man den Anteil der Schüler, die zwischen 130cm und 135cm groß sind, in beiden Verteilungen vergleichen möchte.

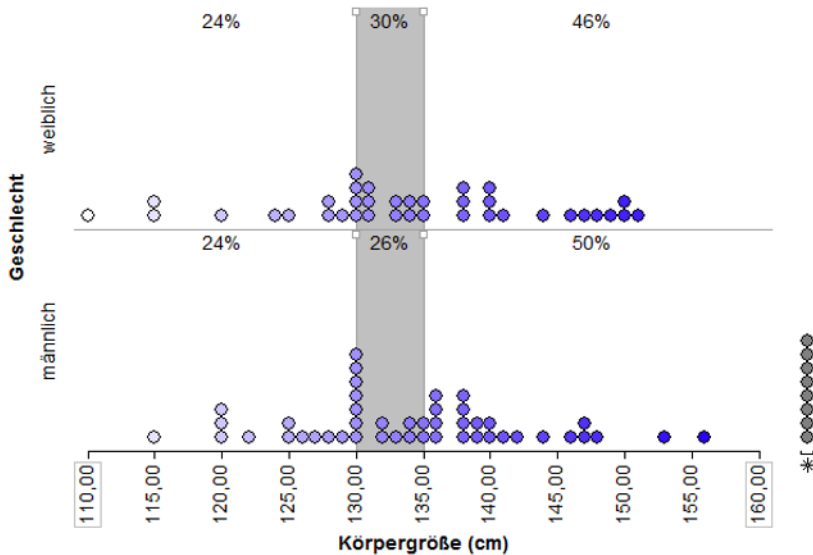


Abbildung 16: Beispiel zum Verteilungsvergleich „Vergleichen gleich großer Klassen anhand von relativen Häufigkeiten“ („klassenbezogen“ bzw. „klassenweise“) (Datensatz aus Dettmar 2013)

In diesem Beispiel würde man konstatieren, dass 30% der Mädchen zwischen 130cm und 135cm groß sind, dieser Anteil bei den Jungen aber nur 26% beträgt. Vergleiche dieser Art, wenn für ein x und ein y (mit $x < y$, aber $x \neq \text{Min}$ und $y \neq \text{Max}$) aus dem Wertebereich die relativen Häufigkeiten $h(x \leq V < y)$ und $h(x \leq W < y)$ verglichen werden, nennen wir „klassenbezogen“²⁸. Diese Vergleiche lassen im Allgemeinen keine tragfähigen Vergleiche bei Verteilungsvergleichen zu.

Wir unterscheiden somit p-basierte Vergleiche, die tragfähig im Rahmen von Verteilungsvergleichen sind und „klassenbezogene“ bzw. „klassenweise“ Vergleiche, die nicht tragfähig im Rahmen von Verteilungsvergleichen sind.

28 In Anlehnung an die Terminologie von Konold et al. (2014) „data as a classifier“.

Vergleich von Verteilungen anhand q -basierter Vergleiche

Die in der Statistik üblichen quantilbasierten Vergleiche, kann man als „Umkehrung“ von p -basierten Vergleichen auffassen. Man gibt feste Prozentsätze vor und ermittelt dann die Stellen in den Datensätzen, bei denen dieser Prozentsatz erreicht ist (siehe Abbildung 17).

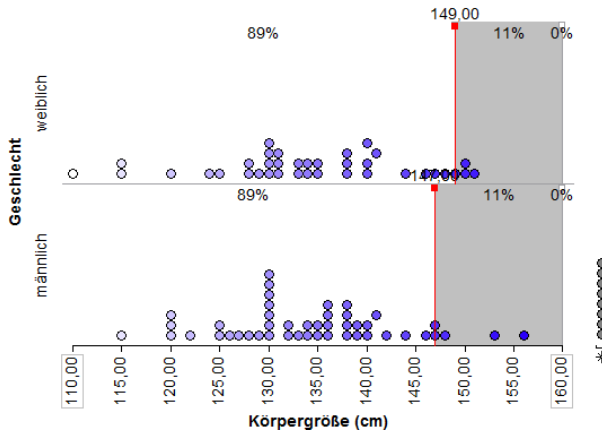


Abbildung 17: Beispiel zum Verteilungsvergleich „Vergleichen gleich großer Anteile“ (Datensatz aus Dettmar 2013)

Biehler nennt diese Vergleichsart „ q -basiert“:

„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, 110)

Speziell beim Boxplotvergleich kann man q -basierte Vergleiche für die beiden Quartile und den Median durchführen.

Zusammenfassend lässt sich sagen, dass es vielfältige Möglichkeiten gibt, Verteilungen zu vergleichen. Wir haben hier nur einige exemplarisch herausgegriffen. Insbesondere haben wir hier nur Vergleiche anhand der einzelnen Verteilungsvergleichselemente (Zentrum, Streuung, etc.) aufgezeigt. In 3.3 und 9.2.1 finden sich auch weiterführende Ausführungen, die es bei der Synthese und Interpretation der Unterschiede zu beachten gilt. So müsste man z.B. in einem weiteren Schritt abwägen, ob eher der Median oder das arithmetische Mittel ein geeignetes Vergleichsmaß (in Bezug auf die Schiefe oder Symmetrie der jeweiligen Verteilungen) darstellt.

Wir halten abschließend folgende Verteilungsvergleich-Elemente fest, die wir als tragfähig erachten:

- Vergleich anhand des Zentrums der jeweiligen Verteilungen
- Vergleich anhand der Streuung der jeweiligen Verteilungen
- Vergleich anhand der Form der jeweiligen Verteilungen
- Herausarbeiten einer Verschiebung zwischen zwei Verteilungen
- P-basierter Vergleich
- Q-basierter Vergleich

Was kann nach dem Vergleichen von Verteilungen eines numerischen Merkmals in der deskriptiven Statistik ein nächster Schritt sein? Wenn es sich um eine Stichprobe aus einer größeren Population handelt, dann kann man fragen, ob bzw. bis zu welchem Grad man die festgestellten Unterschiede auf die Population verallgemeinern kann. In einer praktischen Anwendung könnte sich die Anwendung von Methoden der beurteilenden Statistik (Tests, Konfidenzintervalle) anbieten. Makar und Confrey (2002) unterscheiden hierbei eine Stufe informeller Schlussfolgerungen beim Vergleichen von Verteilungen, die mit der Hinführung zu formalen Verfahren des statistischen Testens fortgesetzt werden kann. Explizit nennen sie als eine wichtige Möglichkeit die Durchführung eines Randomisierungstests. Ähnliche Vorschläge finden sich auch bei Rossman (2008) und Cobb (2007).

2.3.3 Randomisierungstests

Eine Definition für einen Randomisierungstest ist die folgende:

„Ein Randomisierungstest ist ein Permutationstest, der auf einer zufälligen Zuordnung (Randomisierung) der Untersuchungseinheiten (Personen, Beobachtungszeiten, Phasen oder Testblöcke) zu den Behandlungsbedingungen beruht. Die Prüfgröße (*test statistic*) wird nicht nur für die erhaltene Zuordnung, sondern für alle Datenpermutationen berechnet. [...] Unter der Nullhypothese ist das Ergebnis jeder Datenpermutation gleich wahrscheinlich. [...] Falls es möglich ist, die Beobachtungszeiten den Behandlungsbedingungen nach dem Zufall zuzuordnen [...], dann stellen Randomisierungstests für die zufallskritische Auswertung von Einzelfalldaten eine echte Alternative zu den klassischen zeitreihentheoretischen Ansätzen der allgemeinen Klasse der ARIMA (*autoregressive integrated moving average*)-Modelle dar.“²⁹

Eine mathematische Beschreibung des Vorgehens eines Randomisierungstests findet sich bei Ernst (2004). Dabei „prüft [ein Randomisierungstest] die Hypothese eines Gruppenunterschieds, indem man das gefundene Ergebnis mit allen möglichen zufälligen Aufteilungen vergleicht und überprüft, in wie vielen der möglichen Aufteilungen ein gleiches oder noch extremeres Ergebnis aufzufinden ist.“³⁰

29 <https://portal.hogrefe.com/dorsch/randomisierungstest/> (aufgerufen am 13.10.2014)

30 http://www.beltz.de/fileadmin/beltz/downloads/OnlinematerialienPVU/Statistik_und_Forschungsmethoden/09_Kapitel%209_Antworten.pdf (aufgerufen am 13.10.2014)

Die Stichprobe (zufällig vs. nicht zufällig) sowie die Zuordnung der Fälle (die Art der Randomisierung, zufällig vs. nicht zufällig) lässt vier verschiedene Ausgangssituationen auftreten, zwischen denen man hinsichtlich möglicher Schlüsse unterscheiden muss. Diese Szenarien haben Zieffler, Harring und Long (2011) herausgearbeitet und sind in der Tabelle 3 zusammengefasst:

Tabelle 3: "Four Potential Scenarios Researcher Could Face When Making Inferences" - entnommen aus Zieffler et al. (2011, 119)

Scenario	Random sample	Random assignment	Type of Research
1	X		Generalizable research
2		X	Randomized experimental research
3	X	X	Generalizable, randomized experimental research
4			Nongeneralizable, nonexperimental research

Das erste Szenario sieht eine Zufallsstichprobe („Random Sample“) aber keine zufällige Zuordnung („Random assignment“) vor. Der daraus resultierende Typ der Forschung wird nach Zieffler et al. (2011) verallgemeinernd („Generalizable“) genannt, welches zu Schlussfolgerungen hinsichtlich der Population (aus der die Stichprobe gezogen wurde) führt, aber keine kausalen Schlüsse zulässt. Bei Szenario 2, bei dem keine Zufallsstichprobe, dafür aber zufällige Zuordnung gegeben ist, ist es umgekehrt. Laut Zieffler et al. (2011) lässt dieser „Randomized experimental research“ keine Schlussfolgerungen mit Blick auf die Population zu, wohl aber kausale Schlüsse hinsichtlich des Effekts in der Experimental-Gruppe. Ist sowohl eine Zufallsstichprobe als auch eine zufällige Zuordnung entstanden, so liegt „Generalizable, randomized experimental research“ vor. Tritt dieser Fall ein, so können laut Zieffler et al. (2011) sowohl Schlüsse über die Population wie auch Schlüsse zur Kausalität gezogen werden. Aus dem vierten und letzten Szenario, bei dem keine Zufallsstichprobe und keine zufällige Zuordnung vorliegt, lassen sich weder Schlüsse hinsichtlich der Population noch hinsichtlich der Kausalität ziehen. Unabhängig von den in Zieffler et al. (2011) vorgestellten Szenarien, betonen Edgington und Onghena (2007, 6):

„...a randomization test is valid for any kind of sample, regardless of how the sample is selected. This is an extremely important property because the use of non random samples is common in experimentation, and parametric statistical tables (...) are not valid for such samples.“ (Edgington und Onghena 2007, 6)

Cobb (2007) betont die Notwendigkeit der Nutzung einer geeigneten Software beim Durchführen eines solchen Tests. In verschiedenen empirischen Studien, z.B. bei Frischemeier 2013, Frischemeier und Biehler (2014) und Biehler, Frischemeier und Podworny (2015b), haben wir Möglichkeiten aufgezeigt, wie Lernende Randomisierungstests mit TinkerPlots zum einen nach dem oben genannten Szenario 4

(Frischemeier 2013 sowie Frischemeier und Biehler 2014) und zum anderen nach dem oben genannten Szenario 3 (Biehler et al. 2015b) durchführen können.

Im Rahmen dieser Studien haben wir das folgende Schema (siehe Abbildung 18, entnommen aus Biehler et al. 2015b) erstellt, welches einzelne Schritte beim Durchführen eines Randomisierungstests mit TinkerPlots aufzeigt.

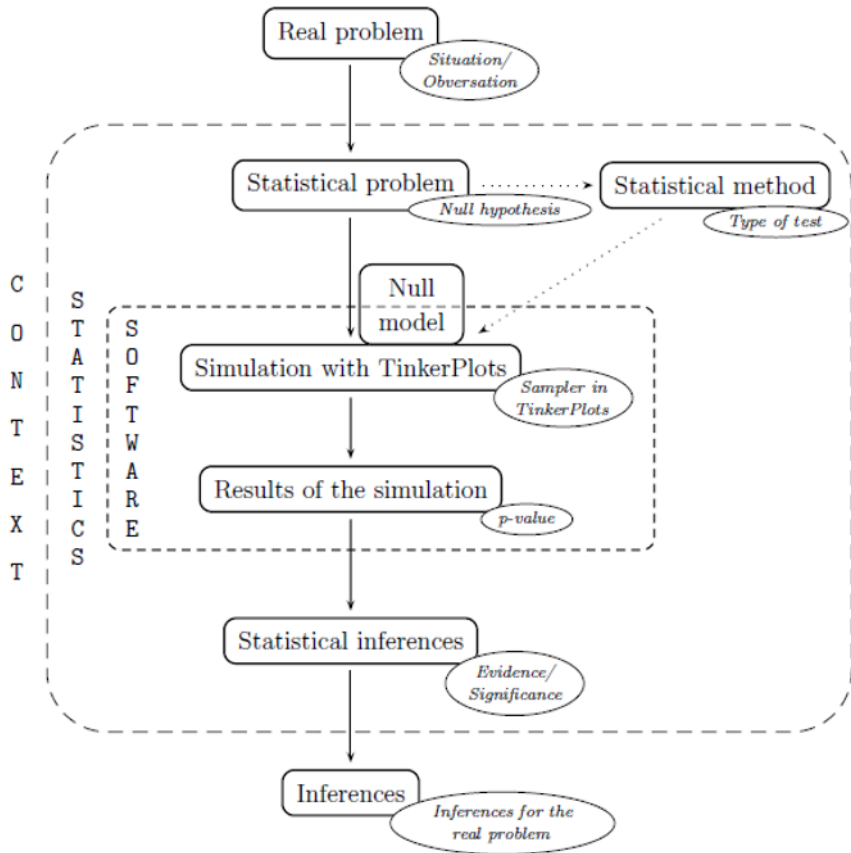


Abbildung 18: Schema: Durchführung einer stochastischen Simulation mit TinkerPlots – entnommen aus Biehler, Frischemeier und Podworny (2015b)

Dieses Schema bezieht sich auf den speziellen Fall von Randomisierungstests in Beobachtungsstudien: Am Anfang steht das reale Problem, was bei einer Aufgabe zum Verteilungsvergleich beispielsweise den Unterschied der arithmetischen Mittelwerte

zweier Verteilungen eines numerischen Merkmals darstellen könnte (wir nennen diese erste Phase daher auch häufig „Beobachtung“). Daraufhin tauchen wir in der zweiten Phase in die statistische Welt ein, in der es hier darum geht, ein statistisches Problem zu formulieren, wie z.B. „sind die Unterschiede zufällig entstanden“? Hier muss nun (siehe Querverbindung) ein statistischer Test (z.B. Randomisierungstest) ausgewählt werden, der zur Beantwortung der Fragestellung beitragen kann. Wählt man einen Randomisierungstest aus, so gilt es nun ein Modell (Null model) zu erstellen und dieses zu simulieren. Man könnte unter Annahme der Hypothese „die Unterschiede zwischen den arithmetischen Mittelwerten sind zufällig entstanden“ nun eine Testgröße (z.B. Unterschiede der arithmetischen Mittelwerte) definieren. Dann könnte man das Modell unter Annahme der Hypothese simulieren, die Referenzverteilung der Testgröße plotten und mit Hilfe des p-Wertes abschätzen, wie groß die Wahrscheinlichkeit ist, dass ein solcher Unterschied oder ein noch größerer Unterschied zwischen den Mittelwerten unter der Annahme besteht, dass die (Null-) Hypothese richtig ist. Nun kann man mit Hilfe des p-Wertes evidenzbasierte Aussagen sowie Rückschlüsse für das reale Problem ziehen. Die Anwendung eines solchen Tests an einem konkreten Beispiel („Geschlechterunterschied bei Schülerinnen und Schülern hinsichtlich ihrer Lesezeit“) mit einer konkreten Software (TinkerPlots) ist im Kapitel 3.4 ausgeführt. Empirische Studien, wie Lernende Randomisierungstests durchführen, sowie unterrichtspraktische Ideen finden sich in Biehler et al. (2003), Frischemeier (2013), Frischemeier und Biehler (2014), Watson (2014) und Biehler et al. 2015b.

Man kann festhalten, dass Randomisierungstests eine gute Möglichkeit bieten an Verteilungsvergleiche anzuschließen sowohl inhaltlich als auch curricular. Auch Randomisierungstests in Szenarien wie dem oben beschriebenen Szenario 4 durchzuführen, kann nicht nur aufgrund des Zitats von Edgington und Onghena (2007, 6) lohnenswert sein. Randomisierungstests, die unter den Bedingungen des oben formulierten Szenarios 4 durchgeführt wurden, lassen zwar keine Verallgemeinerungen aus Befunden (wie z.B. Vergleich von Verteilungen in Beobachtungsstudien) zu. Die Randomisierungstests geben aber die Möglichkeit zu untersuchen, ob festgestellte Unterschiede zwischen den Mittelwerten zweier Verteilungen signifikant größer sind, als wenn die Einteilung in die zwei Gruppen (z.B. Jungen und Mädchen oder Arbeitnehmerinnen und Arbeitnehmer) rein zufällig vorgenommen wurde. Randomisierungstests ermöglichen so den Übergang von informellen zu formellen Schlussfolgerungen (vgl. Harradine, Batanero und Rossman 2011).

2.4 Fundamentale Ideen zur Vermittlung einer Datenkompetenz

In 2.2 wurden verschiedene inhaltsbezogene Anforderungen an Lehrkräfte formuliert. Unter anderem wurden das Durchlaufen eines Datenanalyse-Zyklus, das Arbeiten mit realen Daten, das Betreiben explorativer Datenanalyse sowie das Analysieren von Daten

mit Software aufgegriffen. Diese Ideen sollen im Folgenden genauer charakterisiert werden, um den Bedarf und auch die Einsatzmöglichkeiten dieser in der Lehrer(aus-)bildung zu reflektieren.

2.4.1 Der PPDAC-Zyklus

Wild und Pfannkuch (1999, 226) haben ein vier dimensionales Schema für „statistical thinking in empirical enquiry“ erstellt. Die erste der vier Dimensionen ist der „investigative Cycle“ (PPDAC). Dieser PPDAC-Zyklus umfasst die Phasen *Problem* (Statistische Fragestellungen und Hypothesen generieren), *Plan* (Planen der Datenerhebung), *Data* (Erheben der Daten), *Analysis* (Analyse der Daten) und *Conclusions* (Interpretationen und Schlüsse aus den Daten). Weiterhin differenzieren Wild und Pfannkuch (1999) die Dimensionen „Types of Thinking“ (Dimension 2), „The interrogative Cycle“ (Dimension 3) und „Dispositions“ (Dimension 4). Wir widmen uns vor allem der Dimension 1 („The investigative cycle“, siehe auch Abbildung 19). Die Phase des Problems steht am Anfang und umfasst die Motivation für die Durchführung einer Datenanalyse sowie die genaue Definition des Problems. Ebenfalls sollen bereits auf dieser Stufe verschiedene Einflussfaktoren bedacht und abgewogen werden („grasping system dynamics“). Der zweite Schritt sieht den Plan der Datenerhebung vor. Hier soll das Design der Untersuchung, das Aufstellen und Konstruieren der Messinstrumente sowie konkrete Vorstellungen zum „Data Management“ durchdacht werden. Im Anschluss daran kann dann eine erste Pilotierung erfolgen. Nun folgt die Erhebung der Daten, welche unter anderen das „Verwalten“ der Daten („Data management“) und das Bereinigen der Daten („Data cleaning“) vorsieht. In der Analyse („Analysis“) werden die Daten exploriert, es werden geplante und ungeplante Analysen durchgeführt und Hypothesen generiert.³¹ Am Schluss stehen Schlussfolgerungen („Conclusions“), die die Interpretation der Ergebnisse sowie die Kommunikation weiterer neuer Ideen vorsehen. Dieser Prozess wird auch durch das Schaubild in Abbildung 19 veranschaulicht (Wild und Pfannkuch 1999, 226). Die Durchführung und das eigenständige Erleben dieses Zyklus findet sich fast 1:1 in den Bildungsstandards sowie in den Empfehlungen des AK Stochastik wieder, was diesen Zyklus auch für die Ausbildung unserer Lehramtsstudierenden interessant macht.

Als Elemente statistischen Denkens (Dimension 2: “Types of Thinking”) führen Wild und Pfannkuch (1999, 227) „Recognition of the need for data“, „Transnumeration“, „Variation“, „A distinctive set of models“ und „context knowledge, statistical knowledge and synthesis“ auf.

31 Dass es verschiedene Vorgehensweisen, sowie verschiedenen Typen von “Datenanalytikern” gibt, wird im weiteren Verlauf dieses Kapitels erörtert.

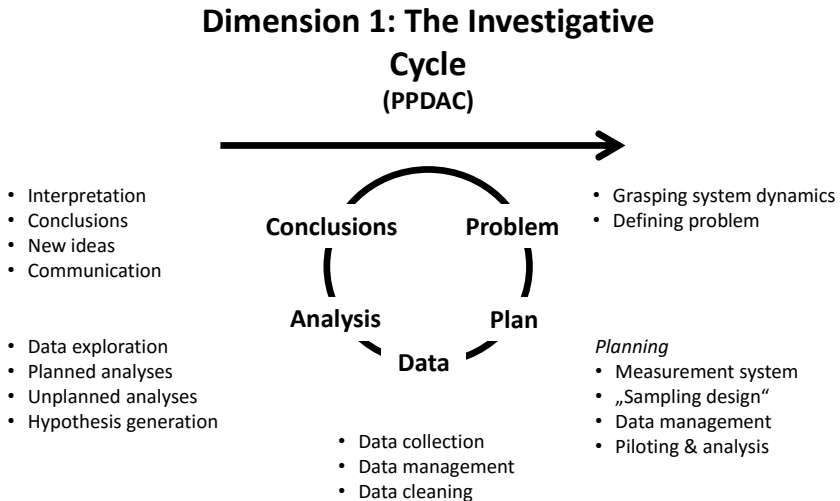


Abbildung 19: The Investigative Cycle (PPDAC) nach Wild und Pfannkuch (1999) – nachgebildet nach Wild und Pfannkuch (1999, 226)

Innerhalb dieser zweiten Dimension etablieren Wild und Pfannkuch den Begriff der „Transnumeration“. Transnumeration bedeutet im übertragenden Sinne eine neue Einsicht in Muster oder Zusammenhänge in den Daten durch das Wechseln der Darstellung oder der Diagrammform. Shaughnessy (2007, 963) beschreibt die Bedeutung von „Transnumeration“ so:

„Wild and Pfannkuch needed a word that went beyond a mere transformation or representation of the data, so as to identify instances in which striking features of context are suddenly revealed. An analogy might be the sudden insight Eureka! Experience that mathematical problem solvers often speak about.“ (Shaughnessy 2007, 963)

Welche Schlüsse lassen sich für die Lehrerbildung ziehen? Shaughnessy (2007) betont in seinem Handbook-Artikel die Wichtigkeit, dass Studierende selbst einen Datenanalyse-Zyklus durchlaufen. Er hebt vor allem die Phasen „Problem“ und „Plan“ hervor, die oftmals vernachlässigt werden (Shaughnessy 2007, 963). Selbiges fordern Burgess (2002), Burgess (2011) sowie Heaton und Mickelson (2002). Letztgenannte betonen, dass Lehramtsanwärter oftmals den Fokus auf den eigentlichen Prozess der Datenanalyse verlieren, weil sie sich vor allem auf die Produktion von Graphiken im Analyseteil konzentrieren. Es stellt sich bei der empirischen Studie von Heaton und Mickelson (2002) heraus, dass Lehramtsanwärter selbst nur unzureichendes Wissen über den Ablauf eines Datenanalyseprozesses haben. So leiten wir als eine wichtige Forde-

rung für die Lehrer(aus-)bildung ab, den PPDAC-Zyklus den Lehramtsstudierenden sowie den Lehrern zugänglich zu machen.

2.4.2 Die explorative Datenanalyse

Maßgeblich zur Entwicklung der explorativen Datenanalyse (kurz: EDA) hat der US-Statistiker John Tukey beigetragen. Er begründete in den 70er Jahren, abgrenzend zur deskriptiven („Beschreiben des Datenmaterials“) und inferenziellen („Testen von Hypothesen“, „Schließen von Stichproben auf Populationen“) Statistik, diese Form der Datenanalyse, welche sich durch eine interaktive und iterative Datenexploration charakterisiert. Es findet ein Prozess statt, in dem Umwege und Irrwege als Lernchancen verstanden werden. Ausgehend von realen Problemen sucht man in den Daten nach Mustern und Besonderheiten und stellt auf Basis dieser Hypothesen auf. Tukey selbst verglich dieses Vorgehen mit der Arbeit eines Detektivs. Ein wesentliches Element der explorativen Datenanalyse (kurz: EDA) ist die Verwendung verschiedener graphischer Darstellungsformen, bei denen man (bzw. der Lernende) versucht Auffälligkeiten zu entdecken und vor dem Hintergrund des Sachproblems zu lösen und zu interpretieren. Darstellungen in der EDA besitzen eine Doppelfunktion: Zum einen repräsentieren sie Daten (simulativ) und zum anderen sind sie ein Mittel für die Tätigkeit des Explorierens (explorativ). Dabei muss der Datendetektiv in der Lage sein, verschiedene Graphiken zu betrachten und zu vergleichen. Aus der Sicht der explorativen Datenanalyse existiert keine einzelne, optimale Graphik. Tukey entwickelte darüber hinaus neue Darstellungsformen wie den Boxplot und das Stängel-Blatt-Diagramm. Fundamental ist das Vergleichen mehrerer Graphiken, welches im Sinne der EDA tiefere Einsichten in das Sachproblem liefert. Anwendungsbeispiele und eine didaktische Analyse dieser EDA-Komponenten finden sich in Biehler (1982).

Aus dieser Vorgehensweise lassen sich auch zwei Typen von Datenanalytikern festmachen: Zum einen die, die den Ideen der explorativen Datenanalyse folgen: Diese tauchen wie ein Detektiv in die Daten ein, fahnden nach Auffälligkeiten und Muster und generieren anhand dieser neue, weiterführende Fragestellungen. Es gibt aber auch Datenanalytiker, die im Sinne der deskriptiven Statistik eher eine zielorientierte Herangehensweise an die Daten verfolgen: Sie haben bereits ein Vorgehensschema (erst Mittelwerte, dann Streuung, ...) verinnerlicht und versuchen dieses sukzessive umzusetzen. Wir machen für unsere Zwecke im Sinne der Unterscheidung eines explorativen oder nicht-explorativen Vorgehens in der Datenanalyse die folgende Charakterisierung: Lernende, die ein ziel-orientiertes Vorgehen verfolgen, verfolgen auf ihrem Weg durch die Daten einen „Wanderweg“. Dieser sieht verschiedene Stationen vor, die nacheinander durchschritten werden. „Ziel-orientierte Lerner“ betrachten die Daten mit einer Theorie, zeichnen sich durch ein zielorientiertes Vorgehen aus und gehen durch die Daten hindurch, um Beweise oder Indizien für die Bestätigung ihrer Theorie zu finden.

Explorativ-orientierte Lernende haben keine spezielle Theorie für sich entwickelt und gehen explorativ durch die Daten, bis ihnen etwas auffällt oder sie ein Muster entdecken. Diese zweite Art des Datenanalytikers ist mit der Charakterisierung der explorativen Datenanalyse sehr kompatibel.³² In diesem Prozess der explorativen Datenanalyse finden sich, wie Pratt, Davies und Connor (2011, 99) ausführen, auch Modellierungsprozesse wieder:

„In EDA, students express their own informal models for the data by searching for trends and patterns in the data, a process often referred to as expressive modelling (Doerr und Pratt 2008).“
(Pratt, Davies und Connor 2011, 99)

Zusammenfassend lässt sich sagen, dass beide Vorgehen (ziel-orientiert vs. explorativ) ihre Berechtigung haben und, dass zukünftige Lehrer beide Vorgehensweisen kennenlernen sollten.

2.4.3 Arbeiten mit realen Datensätzen (im Unterricht und in der Lehrerbildung)

Das Arbeiten mit realen Daten ist fundamental, nicht nur im Stochastikunterricht oder in der Lehrerbildung sondern auch im täglichen Leben. Argumente für ein Arbeiten mit realen Daten (im Mathematikunterricht, aber auch in der Statistik allgemein) finden sich u.a. bei Engel (2007):

„Reale Daten, die von realen Problemsituationen Zeugnis geben, verleihen der Beschäftigung mit Statistik Legitimität und Bedeutung. Sie liefern einen authentischen Grund für die Beschäftigung mit Statistik, warum und wie Daten erhoben wurden und um die Datenanalyse auf einen Kontext zu beziehen“ (Engel 2007, 14)

Borovcnik (2014, 29) betont, dass man „in der modernen Wissens- und Informationsgesellschaft [...] in kaum einen Sektor mehr auf die Verwendung realer Daten verzichten [kann]“ und führt aus, dass

„überall [...] eigens Daten erhoben [werden], um Erkenntnisse zu gewinnen und Entscheidungen zu begründen. Statistische Literalität wird damit zu einem vorrangigen Ausbildungsziel in Statistik.“ (Borovcnik 2014, 29)

Das Arbeiten mit realen Daten im Schulunterricht und in der Lehrerbildung wird von vielen Seiten angeregt, insbesondere von Garfield und Ben-Zvi (2008). Als Argumente dafür nennen sie zum einen die Vielfalt verschiedener Variablen in realen Daten und zum anderen, dass das Potenzial realer Daten durch einen näheren Bezug zur Lebenswelt auch eine höhere Motivation für die jeweiligen Lernenden induziert.

32 Eine Unterscheidung (von Makar und Confrey 2014) zwischen „Wanderer“, „Wonderer“ und „Unwaverer“, anhand der die Unterscheidung zwischen zielorientierten und explorativen Datenanalytikern motiviert wurde, findet sich in 6.1.4.

Diese kann sogar noch erhöht werden, wenn die Daten von den Schülerinnen und Schülern selbst erhoben worden sind. Wie schon in 2.4.1 ausgeführt wurde, gibt es mehrere Möglichkeiten reale Daten zu erheben: durch Befragungen/Umfragen (Umfragen in der Klasse), durch Experimente (z.B. Fallzeit eines Gegenstandes beim freien Fall dokumentieren) und durch Beobachtungen (z.B. Verkehrszählung). Auch im Mathematikunterricht der Grundschule kann diese Art der Datenerhebung schon thematisiert werden (siehe beispielsweise: Biehler und Frischemeier 2013 und Biehler und Frischemeier 2015a). Die Daten können allerdings auch im Internet heruntergeladen werden - Anwendungsbeispiele sowie Ideen zur Umsetzung im Unterricht finden sich in Krüger (2012a) sowie in Biehler und Frischemeier (2015b).

Dass die Thematisierung von realen Daten sowohl im Schulunterricht als auch in der Lehrerbildung nicht trivial ist und einige Fallstricke existieren, heben Pratt et al. (2011, 100) hervor:

„Real data sets present issues that are often not present in sanitised data. For example, difficult numbers, errors in data and missing values are all qualities of data that might be avoided in carefully prepared situations. At some point in a student's education, these issues need to be confronted since they raise important questions about the limitations, scope and reliability of inferences that can be made, as well as techniques for handling the problems.” (Pratt et al. 2011, 100)

Vor allem ist diesbezüglich die Analyse von „Open data“ und „Big data“³³ ein virulentes Thema (siehe Engel 2014). Viele Datensätze sind multivariat (haben eine große Anzahl an Variablen) und laden somit zu vielfältigen Explorationen ein. Krüger (2012a) nutzt Arbeitslosenstatistiken der Bundesagentur für Arbeit und interpretiert diese Daten unter der Frage „Was zeigen die Arbeitslosenzahlen und was verbergen sie?“. Mit Hilfe von Datensätzen der DeStatis-Datenbank des statistischen Bundesamts zeigt sie in einem weiteren Artikel die „Erkundung der Altersverteilung in der Bundesrepublik Deutschland“ auf (Krüger 2012b). Biehler und Frischemeier (2015) bedienen sich ebenfalls der Datenbank des statistischen Bundesamts und untersuchen die Verdienstrukturhebung 2006 auf geschlechterspezifische Unterschiede.³⁴ Möglichkeiten zur Erlangung realer und multivariater Daten ergeben sich einerseits durch die Erhebung einiger Daten (z.B. innerhalb der Schulklasse oder der Schule) oder durch Downloaden von Datensätzen von bekannten Datenbanken, wie zum Beispiel die Datenbank des statistischen Bundesamtes³⁵ oder von „census at school“³⁶. Ein umfassender Überblick über verschiedene Datensätze und Datenbanken findet sich in Engel (2007, 18ff.).

Damit Lehramtsstudierende das Arbeiten mit realen Daten schätzen lernen und zu schätzen wissen, müssen sie selbst mit realen Daten arbeiten und diese explorieren.

33 Für eine Unterscheidung zwischen „Open data“ und „Big data“ siehe Ridgway (2015).

34 Näheres dazu findet sich in Kapitel 9.

35 <https://www.destatis.de/DE/Startseite.html> (aufgerufen am 19.11.2014)

36 <http://new.censusatschool.org.nz/> (aufgerufen am 19.11.2014)

2.4.4 Werkzeugsoftware in der Datenanalyse

Wenn man nun reale Daten explorieren möchte, so ist die Verwendung einer geeigneten Software aus zweierlei Hinsicht unumgänglich: Einerseits ermöglicht sie das Verwalten von großen Datenmengen, andererseits kann sie helfen, Muster und Strukturen in den Daten zu entdecken und wiederzufinden.

„Digital technology facilitates the use of large data sets through its capacity for data storage, easy retrieval and universal availability thanks to the increasing use of idealized data formats.”
(Pratt et al. 2011, 100)

Verschiedene Werkzeuge und eine Übersicht über Technologien für das Lernen und Lehren von Statistik finden sich in Biehler, Ben-Zvi, Bakker und Makar (2013, 650-652). Die Autoren unterscheiden dort zwischen verschiedenen Technologien, die in der Stochastikausbildung in Schule und Hochschule eingesetzt werden können: Statistical software packages (wie SAS, SPSS oder beispielsweise R), Spreadsheets (wie Excel), Applets and stand-alone applications (wie GapMinder), Graphing calculators (wie TI-Nspire), Multimedia materials (wie DataDesk), Data and materials repositories (wie DASL, CAUSE) und Educational software (wie TinkerPlots oder Fathom).³⁷ Während TinkerPlots oder Fathom eher in die Kategorie „educational software“ eingeordnet werden, würde man Excel oder SPSS als „professional software“ bezeichnen.

2.4.4.1 verschiedene Typen von Software

Welche verschiedenen Ansätze gibt es, Software zu charakterisieren? Hier soll kurz auf eine Unterscheidung von zwei Typen eingegangen werden. Zum einen auf die Unterscheidung „Top-down approach“ vs. „bottom-up approach“, zum anderen auf die Unterscheidung zwischen „landscape software“ vs. „route-type software“.

Top-down vs. bottom-up Ansatz

Konold (2006, 6-7) unterscheidet zwei Prinzipien von statistischer Unterrichtssoftware: Zum einen Software, die im Sinne eines „top down“-Ansatzes entwickelt worden ist und zum anderen Software, die nach dem Paradigma eines „bottom-up“-Ansatzes entwickelt wurde. Während eine Software vom Typ „top down“ von „oben herab“ entwickelt wurde, bietet eine Software, die den bottom-up-Approach verfolgt, dem Lerner vielfältige und individuelle Möglichkeiten seinen Lernweg zu gehen (ganz im Sinne des Konstruktivismus). Beide Ansätze werden durch die Abbildung 20 verdeutlicht.

37 Eine ähnliche Stufung (ohne die Aktualität der hier angebrachten Beispiele) findet sich in Chance, Ben-Zvi, Garfield und Medina (2007).

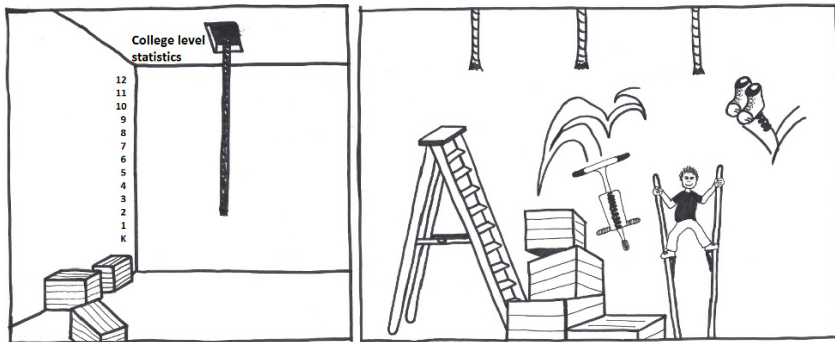


Abbildung 20: Unterscheidung zwischen top-down (links) und bottom-up (rechts) approach – Abbildung nachgebildet (Zeichnung: Svenja Schulz) aus Konold (2006, 6-7)

Die Karikatur in Abbildung 20 verdeutlicht die Philosophie der beiden Software-Typen. Während eine Software von Typ „top-down“ eher von „oben“ nach „unten“ (durch die Klassenstufen hinweg) entwickelt wurde und diesbezüglich ein stringenter Weg verfolgt wird, bietet eine Software vom Typ „bottom-up“ mehrere individuelle Möglichkeiten, Wege zu schaffen und zu erreichen. Von diesen Wegen existiert bei „top-down“-Softwares meistens nur einer, bei „bottom-up“ Softwares existieren oftmals mehrere Lernwege.

Landscape vs. route-type software

Während Konold (2006) eher zwischen einem konstruktivistischen Ansatz und nicht-konstruktivistischen Ansatz unterscheidet, schaut Bakker (2002) auf die Softwaretypologie aus einem anderen Blickwinkel: Bakker (2002) unterscheidet „landscape“ Software vs. „route-type“ Software. Hier wird weniger der Ansatz aus dem Blickpunkt der Lerntheorie reflektiert, sondern vielmehr die Vielfalt und die vielfältigen Wege, die durch die Software gegeben werden. Während eine „route-type“-Software oftmals aus einzelnen Bausteinen (z.B. zum Erstellen eines bestimmten Graphen) besteht, bieten „landscape“-Software ganze Lernumgebungen an. Bei Software vom „route-type“ sind Vorgehen und Lernwege meistens vorbestimmt und weitgehend vorgegeben, weil auch die Funktionalitäten nicht zu viele Möglichkeiten erlauben. Die Schüler können statistische Darstellungen nur nachempfinden aber nicht selbst entdecken. Ein Beispiel ist hier das Softwarepaket „statistical minitools“. Diese werden in Cobb (1999) und Bakker (2004) vorgestellt. Die Lernenden können hier etablierte Darstellungen wie Boxplots oder Histogramme nachempfinden und zwischen diesen switchen. Unter einer Software vom „landscape-type“ versteht Bakker (2002) beispielsweise TinkerPlots. Er räumt diesbezüglich aber auch ein, dass zu viele Möglichkeiten oder zu viele Wege den Lern-

prozess eher behindern als fördern können. So fordert er für diese Art von Software Differenzierungsmöglichkeiten, die es erlauben, gewisse Funktionen ein- und auszuschalten.

Damit sind wir dann schon bei didaktischen Anforderungen an eine Werkzeugsoftware im Bereich der Datenanalyse.

2.4.4.2 didaktische Anforderungen an Werkzeugsoftware in der Datenanalyse

Anforderungen an didaktisch orientierte Softwareprodukte finden sich u.a. in Biehler (1991) und Biehler (1997b). Insgesamt formuliert Biehler (1997b) drei Probleme „complexity of tool problem“, „closed microworld problem“ und „variety problem“, welche in Biehler (1997b, 169-170) diskutiert werden, wobei er unter anderem nützliche Werkzeuge aufführt (Biehler 1997b, 170-171). Als wichtige Punkte nennt er hier, dass eine solche Software explorative Datenanalyse, Simulationen von Zufallsexperimenten sowie statistische Methoden (wie t-Tests, etc.) vereint. Dabei werden unter anderem in Biehler et al. (2013, 649-650) die folgenden Aktivitäten unterschieden:

- „Students can practise graphical and numerical data analysis by developing an exploratory working style
- Students can construct models for random experiments and use computer simulation to study them
- Students can participate in “research in statistics”, that is to say they participate in constructing, analyzing and comparing statistical methods
- Students can use, modify and create “embedded” microworld³⁸ in the software for exploring statistical concepts” (Biehler et al. 2013, 649-650)

Biehler (1997b) fordert darüber hinaus eine einfache Erlernbarkeit der Software, die Vereinigung der Bereiche Datenanalyse, Simulation und Wahrscheinlichkeit, die Möglichkeit bei der Datenanalyse auf unterschiedliche Darstellungen zurückgreifen zu können sowie eine Interaktivität innerhalb der Softwareumgebung. Anforderungen einer Software zum Modellieren und Simulieren finden sich in Biehler (1991, 190).

Fazit

Anhand der Auflistung der Bildungsstandards und der Empfehlungen des AK Stochastik haben wir gesehen, welche Anforderungen an die Schüler gestellt werden. Diese müssen auch von Lehrern im besonderen Maße erfüllt werden. Das Vergleichen von Verteilungen umfasst viele Facetten und bietet das Potenzial -wie wir in 2.3 gesehen haben- vielfältige statistische Konzepte zu nutzen, um Unterschiede zwischen Verteilungen herauszuarbeiten. Diese Aktivitäten sollten mit realen und multivariaten Daten sowie mit Unterstützung geeigneter Software durchgeführt werden. So sollte ein Statistik-Kurs für Lehramtsstudierende Komponenten wie den PPDAC-Zyklus, die Arbeit mit

38 „We use „microworld“ as a notion that comprises exploratory interactive experiments, visualization, and simulations, and applets.“ (Biehler et al. 2013, 650)

realen und multivariaten Daten, sowie den Einsatz adäquater Software enthalten. Dabei soll im Folgenden ganz besonders auf die „adäquate Software“ eingegangen und TinkerPlots als Medium des Lehrers, des Lehramtsstudierenden und des Schülers vorgestellt werden.

In Kapitel 3 soll ausgeführt werden, dass TinkerPlots diesen Ansprüchen genügt. Dabei wird die Software zunächst kurz allgemein vorgestellt und dann unter verschiedenen Gesichtspunkten beleuchtet: Als „educational software“, die das Erlernen von Datenanalyse fördert, als „Werkzeug“ zur Datenanalyse, die die Exploration komplexer Datensätze erlaubt und als Tool für weitere Experimente, wie z.B. das Durchführen eines Randomisierungstests (siehe Kapitel 2.3.3).

Statistisch denken und forschen lernen mit der
Software TinkerPlots

Frischemeier, D.

2017, XXIV, 654 S. 344 Abb., Softcover

ISBN: 978-3-658-15322-9