Eine empirische Studie über die Kompetenz von Schülerinnen und Schülern der Sekundarstufe I im Umgang mit bivariaten Daten bei verschiedenen Darstellungsformen

Inauguraldissertation zur Erlangung der Doktorwürde (Dr. phil.)
der Fakultät für Mathematik der Universität Bielefeld

vorgelegt von
Watzl Christoph
aus
93051 Regensburg

Bielefeld, im Juni 2018

Inhaltsverzeichnis

Dank(barkeit)

Was lange währt, wird endlich gut. Nach so langer Zeit des sprichwörtlichen „Herumdokterns“ an meiner Dissertation gilt mein Dank besonders meinem Doktorvater Prof. Dr. Michael Kleine, der mir in mehrfacher Hinsicht den nötigen Freiraum entgegengebracht hat, meine Arbeit trotz meiner vielfältigen Betätigungsfelder (als Seminarrektor und Lehrkraft mit vollem Deputat, als Buchautor, als Lehrbeauftragter der Universität Regensburg und – last but not least – als mittlerweile dreifacher Familienvater) zu vollenden. Seine Geduld, seine Tipps und seine Begeisterung für meine Arbeit, in die er sich nicht selten ganz von vorne eindenken musste, waren eine der wichtigsten Säulen für das Gelingen dieses Projekts. Selbstverständlich gebührt auch allen Lehrkräften, Schülerinnen und Schülern, die an der Durchführung der Studie beteiligt waren, ganz besonderer Dank.

Von Herzen danke ich meiner Frau Heidi, die mir in all der Zeit des Recherchierens, Denkens und Schreibens stets den Rücken freigehalten hat und sich immer unserer Kinder angenommen hat, wenn Papi wieder einmal „im Büro arbeiten“ musste. Ohne ihr Organisationstalent, ihre Geduld und ihr Verständnis hätte ich diese Arbeit wohl nicht vollenden können.

1 Einleitung: Warum bivariate Daten?

„Daten und Zufall“ – auch für mich als junge Mathematiklehrkraft kam *damals* im Schuljahr 2007/08 dieses *neue* Lehrplankapitel quasi aus dem Nichts. Mit der Einführung eines neuen Lehrplans an den bayerischen Realschulen wurden die Beschlüsse der Kultusministerkonferenz von 2003 umgesetzt, so dass fortan in fast jeder Jahrgangsstufe eine Unterrichtssequenz zu diesem Themengebiet verbindlich zu behandeln war.

Wie nicht wenig anderen Lehrkräften standen mir damals viele Fragezeichen ins Gesicht geschrieben, als ich mich daran machte, die ersten Unterrichtsstunden für meine Klassen zu planen. Zudem fühlte ich mich verpflichtet, meiner Vorbildfunktion als Seminarlehrkraft gerecht zu werden und meinen Referendar/inn/en aufzuzeigen, wie *guter* Statistik- und Stochastikunterricht aussehen könnte. Also habe ich mich – vielleicht als ohnehin tendenziell statistikaffine Persönlichkeit – in das neue Themenfeld eingelesen, recherchiert und fortgebildet – und die Datenanalyse, besonders das Suchen und Entdecken von Mustern in Daten, für mich entdeckt. Ermutigt durch didaktisch wertvolle Beiträge und Beispiele in der Literatur, aber zugleich enttäuscht von der teilweise *vorsichtigen Aversion* vieler Lehrkräfte, Lehramtsanwärter und Studenten gegenüber diesem Themengebiet, war und ist es umso mehr mein Bestreben, es den Schüler/inne/n und Junglehrern so spannend, interessant und nützlich wie möglich nahe zu bringen. Besonders nach der Lektüre des Buches „Daten und Zufall“ von Eichler und Vogel wage ich mich im Unterricht von Mal zu Mal selbstbewusster an das Thema Statistik und bivariate Datensätze heran – gleichwohl bivariate Daten nach wie vor im Lehrplan (und auch allen anderen Lehr- und Bildungsplänen der Republik) keine explizite Benennung finden.

Durch die Beschäftigung mit dieser Thematik ergaben sich letztendlich auch einige der in dieser Arbeit untersuchten Fragestellungen: Ist die Beschäftigung mit bivariaten Daten wirklich schon in den unteren Klassen der Sek. I möglich? Erkennen Schüler/innen in einer 6. Jahrgangsstufe wirklich einen Trend in den Daten, wenn sie ein Diagramm oder eine Tabelle intensiver betrachten? Und können sie diesen entsprechend verbalisieren?

Darüber hinaus habe ich mich oft gefragt, warum das Thema *Daten und Zufall* immer noch so unbeliebt erscheint: Auch Jahre später, nach denen der Themenkomplex *Daten und Zufall* vermeintlich zum Alltagsgeschäft der Lehrenden gehört, herrscht auf diesem Gebiet bei vielen Lehrkräften immer noch große Unsicherheit bzw. *didaktische Hilflosigkeit* – und die meisten Unterrichtenden machen auch keinen Hehl daraus. Sie wagen sich nur zögerlich an dieses Thema, weil Datenanalyse und Stochastik entweder überhaupt nicht Bestandteil ihres Studiums oder im Gegenteil maßtheoretisch völlig überfrachtet war (vgl. EICHLER & VOGEL, S. X). In einigen Bundesländern war es „zum Teil Jahrzehnte" (EICHLER & VOGEL, Vorwort) lang überhaupt kein Bestandteil curricularer Vorgaben.

Damit der Aufbau einer Datenkompetenz aber bei Schüler/inne/n gelingen kann, ist Motivation und Interesse am Thema Daten und Zufall unbedingte Voraussetzung. Die bereits erwähnte und immer noch anzutreffende reservierte Haltung etlicher Lehrkräfte – und Schüler/innen! – gegenüber dem Thema Daten und Zufall ist jedoch nach wie vor im Unterrichtsalltag beobachtbar. Die folgenden drei Punkte scheinen mitverantwortlich zu sein, dass sich die Ressentiments, insbesondere gegenüber dem Bereich der Datenanalyse und Statistik, weiterhin halten:

1. Nach Einführung der üblichen statistischen Kenngrößen (relative Häufigkeit, arithmetisches Mittel, Median, Spannweite, ...) verkommt der Unterricht häufig zur bloßen Berechnung dieser Werte; eine Vorgehensweise, die sich vor allem in den Jahrgangsstufen 5 und 6 mitunter deswegen als mühsam erweisen kann, da in der Regel kein Taschenrechner zur Verfügung steht. Schulbücher, die um diese Widrigkeit wissen, verwenden aus diesem Grund oft *geschöntes* Zahlenmaterial, um die Berechnungen halbwegs praktikabel zu halten. Dies allerdings läuft der Authentizität von Daten zuwider: *„Use real data!"* ist die eine der wesentlichsten Forderungen für den Aufbau statistischer Kompetenz (vgl. v. a. GARFIELD & BEN-ZVI 2008, p. 14 f).

2. Die errechneten Kenngrößen werden nur in seltenen Fällen für eine anschließende Analyse des Datenmaterials bzw. zur Interpretation des Sachverhalts verwendet; ein Vergleich dieser Werte und ein Hinterfragen der (jeweiligen) Aussagekraft fehlt häufig (vgl. EICHLER & VOGEL 2013, S. 56). Vorschnell wird nach der Korrektur des berechneten Ergebnisses, das auf diese Art lediglich einen wenig sinnhaften und isolierten Wert darstellt, möglichst rasch zur nächsten Aufgabe übergegangen, um das *Einschleifen* der Rechentechnik voranzutreiben.

3. Meist ist die Ursache mangelnder Motivation bereits in der Phase noch *vor* dem *Rechnen* mit statistischen Kenngrößen begründet: In der Regel wird im Unterricht auf vorgefertigte Datensätze und Beispielaufgaben in Schulbüchern zurückgegriffen, die verständlicherweise nicht dieselbe Begeisterung auslösen wie selbst erhobene Daten. Alleine die Thematik *Daten erheben* birgt bei gut gemeinter, aber unwissender Durchführung viele didaktische Tücken, so dass sich diese „vermeintlich attraktive Handlungsorientierung“ ohne „kontextuell motivierten Beweggründe“ als „trügerisch und mathematisch-inhaltlich hohl“ erweist (EICHLER &VOGEL 2013, S. 119). Datenerhebung, wie in den KMK-Beschlüssen und den Bildungsplänen verbindlich verankert, muss im Unterricht noch deutlich mehr Raum zugestanden werden: „Die Datenanalyse basiert stets auf dem fragenden Interesse der Schülerinnen und Schüler. [...] Fragen zu realen Phänomenen sind der Ausgangspunkt, die Motivation und der stete Kontrollmechanismus allen Arbeitens mit den Daten.“ (EICHLER & VOGEL 2013, S. 119)

Die Erweiterung auf bivariate Datensätze scheint somit mehr als ein probates Mittel zu sein, Datenanalyse motivierender und zugleich authentischer zu gestalten. EICHLER und VOGEL (2013, S. 75) plädieren allein deswegen schon für die unbedingte Thematisierung der bivariaten Datenanalyse in der Sekundarstufe I, „da sonst die Datenanalyse [...] um einen ihrer spannendsten Aspekte reduziert würde.“

Mit der vorliegenden Arbeit möchte im anschließenden Theorieteil zuerst aufzeigen, wie sich der Umgang mit bivariaten Daten bislang im Unterricht der Sekundarstufe darstellt bzw. darstellen kann, wie bivariate Datenanalyse *funktioniert* und welche Erfahrungen es bereits dazu gibt. Im empirischen Teil soll dann anhand von sechs Forschungsfragen untersucht werden, wie datenkompetent Realschüler/innen aus jeweils drei 6. und 9. Klassen im Umgang mit bivariaten Daten sind.

2 Theoretischer Rahmen

Zusammenfassung:

Nach einer ersten Annäherung des Begriffs der (allgemeinen) Datenkompetenz stellt sich die Frage, inwieweit sich diese Überlegungen auf bivariate Daten beziehen. Dazu wird der mathematische und didaktische Hintergrund beleuchtet und an einem bivariaten Datensatz beispielhaft illustriert. Nach einem Blick in die Bildungsstandards und Lehrpläne wird im Anschluss unter Berücksichtigung aller bis dahin beschriebenen Aspekte ein hierarchisches Modell des Konstrukts Datenkompetenz im Umgang mit bivariaten Daten *gefasst, das als Grundlage für die durchgeführte Studie dienen soll. Das abschließende Kapitel gibt einen Überblick über den aktuellen Forschungsstand.*

2.1 (Allgemeine) Datenkompetenz: Ein Begriff, viele Facetten

> „Es ist überwältigend, welche Rolle Daten bei Entscheidungen in der Geschäftswelt, der Politik, der Forschung und im täglichen Leben spielen. Konsumentenumfragen bestimmen die Entwicklung und das Marketing neuer Produkte. Meinungsumfragen bilden die Grundlagen von Strategien politischer Kampagnen, und Experimente werden eingesetzt, um die Sicherheit und Wirksamkeit neuer medizinischer Behandlungsmethoden zu bewerten. Statistiken werden oft auch missbraucht, um die öffentliche Meinung zu beeinflussen oder um die Qualität und Effektivität kommerzieller Produkte fälschlich darzustellen. Schülerinnen und Schüler brauchen Grundkenntnisse von Datenanalyse und der Wahrscheinlichkeitsrechnung, um statistisch argumentieren zu können – Fertigkeiten, die für informierte Staatsbürger und intelligente Konsumenten notwendig sind.“ (NCTM, 2000, dt. Übersetzung nach Bescherer C. & Engel J.)

Mit diesen Worten hat das NCTM (= *National Council of Teachers of Mathematics*; Sitz in Virginia/USA), die weltweit größte Organisation, die sich mit mathematischer Bildung befasst, bereits vor knapp 20 Jahren die ungemeine Bedeutung statistischer Bildung treffend begründet. Doch wie lassen sich diese Grundkenntnisse und Fertigkeiten beschreiben? Was bedeutet es, „statistisch argumentieren“ zu können?

In der Regel fällt bei Fragen und Diskussionen rund um statistische Grundkenntnisse schnell der Begriff Datenkompetenz, dessen Definition sich als typisches *Kompetenz-Dilemma* (vgl. WATSON 2002, p. 26) erweist: Datenkompetenz ist ähnlich wie Medienkompetenz oder soziale Kompetenz schwierig zu beschreiben, denn von „Kompetenz spricht man offensichtlich immer dann, wenn man vor dem Dilemma steht, dass man eine ganz Fülle von Fertigkeiten, Fähigkeiten, Begriffen, Einsichten etc. meint, sich aber kaum in der Lage sieht, dieses alles allumfassend zu beschrieben“ (BIEHLER & WEBER 1995, S. 5). Und mit exakt dieser Formulierung versuchen BIEHLER und WEBER dann direkt im darauffolgenden Absatz, Datenkompetenz zu definieren: Eben als genau diese „Fülle von Fertigkeiten, Fähigkeiten, Begriffen, Einsichten etc., die notwendig oder zumindest hilfreich beim sachgerechten Umgang mit Daten sind.“ (BIEHLER & WEBER 1995, S. 5).

So finden sich in der Literatur in der Regel auch nur Umschreibungen oder Aufzählungen einzelner Teilkompetenzen, die versuchen, diese vielen Facetten von Datenkompetenz zu konkretisieren. WAGNER hat in Anlehnung an die „*Bestandteile* der Datenkompetenz“ nach HANCOCK (1995, S. 33) und der Aufschlüsselung nach BIEHLER (2001, S. 98) versucht, Datenkompetenz zusammenfassend auf die folgenden elf Punkte einzugrenzen bzw. zu konkretisieren. Demnach umfasst Datenkompetenz (vgl. WAGNER 2006, S. 18 f):

1) Methoden der Datenerhebung kennen und eine solche durchführen können
2) Daten je nach Bedarf und Zweck anordnen und strukturieren können
3) Verschiedenartige grafische Darstellungen von Daten lesen und interpretieren können
4) Statistische Kenngrößen nutzen und interpretieren können
5) Fähigkeit, ggf. von der Stichprobe auf die Population zu schließen
6) Planung einer Datenanalyse durch entsprechende Fragestellung
7) Methoden zur Datenanalyse auswählen und reflektieren
8) Interpretation von Darstellungen und Ergebnissen im Sachkontext
9) Aussagekraft von untersuchten Daten beurteilen
10) Kommunikation über die Resultate einer Analyse
11) Umgang mit entsprechender Software zur Datenanalyse

Auch im englischsprachigen Raum, in der die fachdidaktische Diskussion um Statistik auf eine längere Tradition zurückblicken kann (v. a. in den USA, Australien und Neuseeland), stößt man auf die ähnliche Problematik, dass sich *Data Literacy* bzw. *Statistical Literacy* – diese beiden Begriffe werden synonym für den deutschsprachigen Begriff Datenkompetenz benutzt – nicht eindeutig benennen bzw. definieren lassen. Zwar haben sich dort die drei „Konstrukte" *Statistical Reasoning*, *Statistical Thinking* und *Statistical Literacy* etabliert, aber auch hier „ist eine genaue Grenzziehung schwierig. Teilweise werden die einzelnen Begriffe uneinheitlich verwendet, teilweise werden sie als gegeneinander austauschbar angesehen (Bidgood, 2014; Chance, 2002; delMas, 2002)" (SPROESSER 2015, S. 9).

- ***Statistical literacy*** includes basic and important skills that may be used in understanding statistical information or research results. These skills include being able to organize data, construct and display tables, and work with different representations of data. Statistical literacy also includes an understanding of concepts, vocabulary, and symbols, and includes an understanding of probability as a measure of uncertainty.
- ***Statistical reasoning*** may be defined as the way people reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, representations of data, or statistical summaries of data. Statistical reasoning may involve connecting one concept to another (e.g., center and spread), or it may combine ideas about data and chance. Reasoning means understanding and being able to explain statistical processes and being able to fully interpret statistical results.
- ***Statistical thinking*** involves an understanding of why and how statistical investigations are conducted and the "big ideas" that underlie statistical investigations. These ideas include the omnipresent nature of variation and when and how to use appropriate methods of data analysis such as numerical summaries and visual displays of data. Statistical thinking involves an understanding of the nature of sampling, how we make inferences from samples to populations, and why designed experiments are needed in order to establish causation. It includes an understanding of how models are used to simulate random phenomena, how data are produced to estimate probabilities, and how, when, and why existing inferential tools can be used to aid an investigative process. Statistical thinking also includes being able to understand and utilize the context of a problem in forming investigations and drawing conclusions, and recognizing and understanding the entire process (from question posing to data collection to choosing analyses to testing assumptions, etc.). Finally, statistical thinkers are able to critique and evaluate results of a problem solved or a statistical study.

Abb. 2-1: *„Although no formal agreement has been made …"*: Der Versuch einer Zusammenfassung der unterschiedlichen Auffassungen zu *Statistical literacy*, *Statistical reasoning* und *Statistical thinking*. (BEN-ZVI & GARFIELD 2005, S. 7)

Bei der Durchsicht der durchaus vielfältigen Literatur zu diesen drei Begrifflichkeiten scheint es zudem so, als ob sich die Diskussion um diese Konstrukte ein wenig verselbstständigt hätte; sie sind für einen Definitionsversuch bzw. Begriffseingrenzung von *Data Literacy* bzw. *Stastistical Literacy* nur bedingt hilfreich. Einige Autoren sehen in der zusammenfassenden Darstellung von Ben-Zvi und Garfield (vgl. Abb. 2-1 auf S. 9) ein hierarchisches Modell, andere wiederum fassen *Statistical Literacy* als Oberbegriff auf, der *Statistical Reasoning* und *Statistical Thinking* miteinschließt (vgl. Sproesser 2015, S. 19).
Sehr häufig wird bei der Diskussion um diese drei Konstrukte auf die so genannten fünf Säulen statistischen Denkens („*the foundations, on which statistical thinking rests*“) nach Wild und Pfannkuch (1999, p. 227 f) verwiesen, die in fast allen (auch deutschsprachigen) Abhandlungen als fundamental im Zusammenhang mit Datenkompetenz angesehen werden. Diese lauten (mit entsprechender kurzer Erläuterung, vgl. dazu Eichler & Vogel 2013, XII und Vogel & Eichler 2010b, S. 882):

• *Recognition of the need for data:*	Einsicht in die Notwendigkeit von Daten, die die Voraussetzung für einen fundierten Erkenntnisgewinn darstellen.
• *Transnumeration:*	Fähigkeit der flexiblen Datenaufbereitung: Je nach Perspektive bzw. Problemstellung ist u. U. eine andere Darstellung hilfreicher
• *Variation:*	Einsicht in die Variabilität statistischer Daten: Die Ergebnisse statistischer Erhebungen lassen sich prinzipiell nicht genau vorhersehen, selbst bei Erhebungen unter denselben Rahmenbedingungen.
• *A distinctive set of models:*	Datenanalyse besteht im Suchen, Identifizieren und Beschreiben von Mustern in den Daten. Dabei folgt sie ihren eigenen, typischen Modellierungsansätzen.
• *Context knowledge, statistical knowledge and synthesis:*	Statistik und realer Kontext stehen in einem sich gegenseitig bedingenden Verhältnis.

Neben den bisherigen Begriffseingrenzungen und -abgrenzungen, Umschreibungen bzw. der Konstatierung wesentlicher Säulen von Datenkompetenz wird in der fachdidaktischen Diskussion häufig auch auf das Modell von Gal (2002) Bezug genommen, das vor allem in Hinblick auf das *Lesen* grafischer Darstellungen

(„*reading contexts*“) die Schlüsselqualifikationen („*key components*“: GAL 2002, p. 19) von *Statistical literacy* zu vereinen sucht (Abb. 2-2).

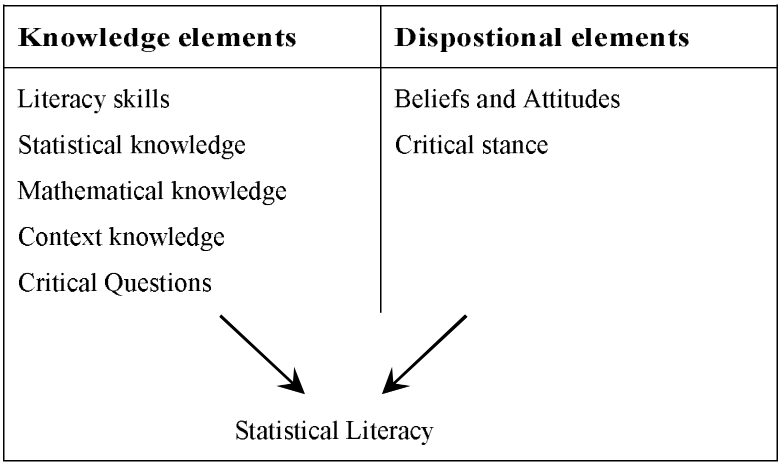


Abb. 2-2: „*A model of statistical literacy*“ (GAL 2002, p. 4)

Demnach setzt sich *Statistical literacy* aus Wissens- und Wesenskomponenten zusammen. Dazu gehören

> „… ‚Literacy Skills‘, also allgemeine Kenntnisse und Fähigkeiten des Entnehmens und Verarbeitens von Information aus schriftlichen Darstellungen wie Listen, Tabellen oder Grafiken […]. Daneben ist spezifisches Wissen im Bereich der Statistik (z. B. grundlegende Begriffe bezüglich deskriptiver Statistiken oder Darstellungen von Daten), der Mathematik (z. B. mathematische Prozeduren zur Bestimmung von Kennwerten wie dem Mittelwert oder dem Median) und des jeweiligen Datenkontexts (z. B. ob alle relevanten Informationen eines Studiendesigns dokumentiert sind, ob statistische Schlüsse auf Grundlage dieses Designs zulässig sind, …) von Bedeutung. Durch das separate Aufführen von sogenannten kritischen Fähigkeiten und einer kritischen Haltung betont Gal, dass über eine grundsätzlich kritische Einstellung gegenüber statistischen Informationen und Schlussfolgerungen hinaus spezielle Kenntnisse beispielsweise über mögliche Verzerrungen durch das Verwenden bestimmter Kennwerte oder Darstellungen erforderlich sind. Um diese Elemente mit dem Ziel eines adäquaten Umgangs mit Daten zu verbinden, sind motivationale Variablen wie die positive Einschätzung der eigenen Fähigkeiten sowie das Interesse und die

> Wertschätzung gegenüber statistischer Verfahren notwendig, die somit das letzte Element von Gal's Modell bilden." (SPROESSER 2015, S. 17)

Die in diesem Modell geforderte kritische Grundhaltung wird im Übrigen beinahe von allen Autoren als einer der bedeutendsten Bestandteile von Datenkompetenz angesehen. Sie hat allerdings nicht ein (manchmal falschverstandenes) grundsätzliches Misstrauen gegenüber jeglicher statistischen Darstellung oder gar der Statistik als Wissenschaft allgemein zum Ziel, sondern meint vielmehr ein „Sensibilisieren" (vgl. BIEHLER, 2006, S. 114), eine „kritische[...] Haltung gegenüber datengestützter Argumentation anderer", die Fähigkeit, „Zahlen und Graphiken nicht blind zu vertrauen, sondern diese kritisch zu durchleuchten" (WAGNER 2006, S. 20). Datenkompetente Bürger sollten in der Lage sein, „datenbasierte Entscheidungen" zu treffen und „auf Daten basierende Fehler [und] Halbwahrheiten [...] als solche zu identifizieren." (SPROESSER 2015, S. 17)

Die in diesem Kapitel „aufgezeigte fehlende Einheitlichkeit an Konstruktbeschreibungen und -abgrenzungen macht deutlich, warum einige Wissenschaftler für theoretische und empirische Arbeiten in diesem Bereich empfehlen, zunächst eine klare Arbeitsdefinition des zu untersuchenden Konstrukts vorzulegen (z. B. Ben-Zvi & Garfield, 2005; Shaughnessy, 2007)" (SPROESSER 2015, S. 19). Bevor allerdings dieser Empfehlung entsprochen wird (→ Kap. 2.4) und erläutert wird, welches Verständnis von Datenkompetenz dieser Studie zu Grunde liegt, sollen in den nächsten beiden Kapiteln zuerst notwendige fachliche und anschließend didaktische Aspekte im Umgang mit *bivariaten* Daten beleuchtet werden.

2.2 Sachanalyse: Begriffe und Datenanalyse bivariater Statistik

Die bisherigen Ausführungen machen die Vielschichtigkeit des Begriffs Datenkompetenz deutlich, lassen aber eine vor allem für diese Arbeit entscheidende Frage offen: Inwieweit beziehen sich die herausgestellten Aspekte von Datenkompetenz auf den Umgang mit *bivariaten* Daten?

Die Antwort auf diese Frage liegt in der Struktur bzw. Beschaffenheit der erhobenen bzw. vorliegenden Daten selbst, denn bei der Betrachtung bivariater Daten verändert sich in aller Regel die Sichtweise: im Mittelpunkt steht dann vor allem die Frage nach Zusammenhängen in den Daten. Begriffe wie Streudiagramm, Korrelation, Regression und Kausalität drängen sich in den Vordergrund. Die Analyse univariater Daten kann „zwar auch interessierende Fragen beantworten, sobald aber im Datenmaterial mehr als zwei Merkmale enthalten sind, besteht ein unwillkürliches Interesse an Zusammenhängen." (EICHLER & VOGEL 2013, S. 75). Die Fragen, die sich bei einer bivariaten Datenanalyse sodann stellen, zielen „fast zwangsläufig auf einen Vergleich oder auf die Suche nach Ursache und Wirkung." (EICHLER & VOGEL 2013, S. 119).

Bereits aus rein ökonomischen Gründen werden bei einer Datenerhebung für gewöhnlich mehrere Merkmale erfasst. Auch wenn es sich gegebenenfalls nur um eine kleine Erhebung handelt, ist grundsätzlich auf die Qualität der Erhebung zu achten:

> „Wenn man [...] erkannt hat, dass man für Antworten auf viele Fragen statistische Daten benötigt, dann stellt sich die Frage, wie die Daten beschaffen sein müssen, um möglichst gute Antworten zu finden. [...] Das heißt, dass nicht die Größe einer Stichprobe primär wichtig ist, sondern die Beschaffenheit der Stichprobe, die sorgfältige Definition von Merkmalen und die ebenso sorgfältige Erhebung bzw. Messung von Merkmalsausprägungen." (EICHLER & VOGEL 2013, S. 16).

Das theoretische Wissensfundamentum zu den eben erwähnten Begrifflichkeiten und Möglichkeiten der Datenauswertung, ebenso aber zu den dafür notwendigen Voraussetzungen wie z. B. die Mitberücksichtigung der Merkmalsskalen, soll nun in den folgenden Unterkapiteln erläutert werden.

2.2.1 Merkmale, Skalen und weitere grundlegende Begriffe

„Der Begriff **Merkmal** ist ein Grundbegriff der Beschreibenden Statistik" und kann als „Eigenschaft von Personen oder Objekten, die bei einer statistischen Untersuchung von Interesse sind, erklärt werden." (KRÜGER, SILL und SIKORA 2015, S. 29). Die dabei untersuchten Personen oder Objekte werden als Träger dieser

Eigenschaften demzufolge auch als **Merkmalsträger** oder statistische Einheit bezeichnet. Die möglichen Werte oder Kategorien, die ein Merkmal dabei annehmen kann, werden **Merkmalsausprägung**en genannt. Diese Merkmalsausprägungen können üblicherweise eine der folgenden drei Kategorien (**Skalen**) zugeordnet werden:

- **Nominalskala**: Die Merkmalsausprägungen genügen keiner Hierarchie, d. h. sie können zwar *benannt* und unterschieden werden, nehmen aber keinen mathematischen Wert an (z. B. Geschlecht, Konfession, Staatsbürgerschaft, Beruf, Sorte, ...).
- **Ordinalskala** (Rangskala): Die Merkmalsausprägungen können zwar in eine Reihenfolge gebracht werden, allerdings sind in dieser Hierarchie die Abstände zwischen den Merkmalsausprägungen nicht definiert (z. B. Zensuren, Güteklassen, Rangplätze, ...).
- **Kardinalskala** (metrische Skala): Die geordneten Skalenwerte können auch hinsichtlich ihres Abstandes verglichen werden (z. B. Alter, Währung, Gewicht, Längen, Zeit, ...). (intervallskaliert)

Die Abbildung 2-3 (S. 15) stellt diese Begriffe exemplarisch noch einmal in anschaulicher Weise dar.

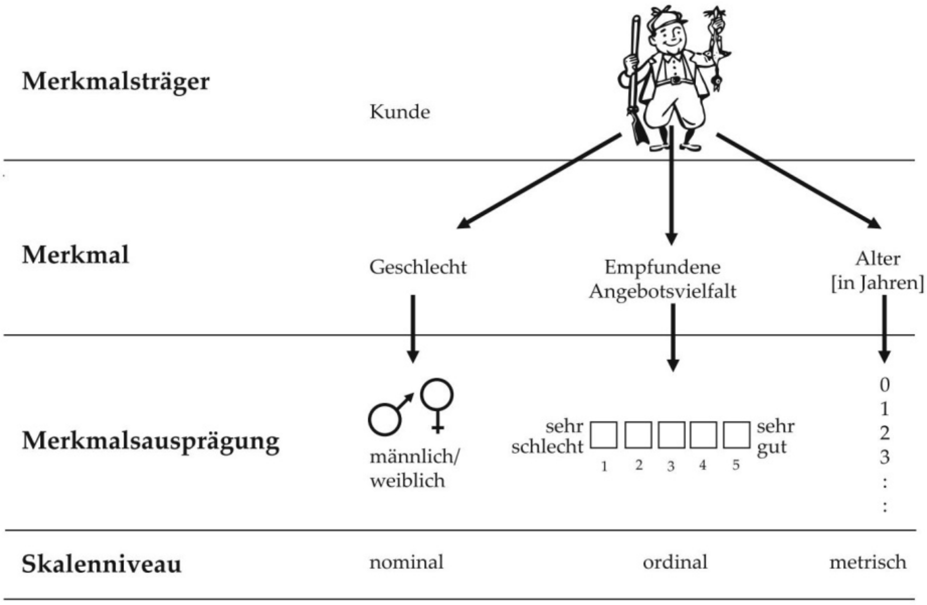


Abb. 2-3: Merkmalsträger/Merkmale/Merkmalsausprägung/Skalenniveau (CLEFF 2015, S. 19)

Daten, die sich lediglich auf ein Merkmal beziehen, heißen **univariat**. Daten, die sich auf zwei Merkmale beziehen, heißen **bivariat**. Bei der Beobachtung von drei oder gar noch mehr Merkmalen, spricht man von **multivariaten** Daten (vgl. EICHLER & VOGEL 2013, S. 19).

Da der Fokus dieser Arbeit auf die Analyse bivariater Daten zielt, wird im Folgenden auf die mathematische Erläuterung der klassischen Kenngrößen einer eindimensionalen (univariaten) Analyse verzichtet bzw. an dieser Stelle auf einschlägige Literatur verwiesen (z. B. CRAMER & KAMPS 2014). Der Vollständigkeit halber und für einen besseren Überblick sind sie zumindest stichpunktartig in der Abbildung 2-4 (S. 16) dargestellt.

Auch wenn nach einer Umfrage in der Regel mehrere Merkmale samt deren Ausprägungen zur Auswertung vorliegen, sind manchmal durchaus Aussagen über ein einziges Merkmal von Interesse, die dann mithilfe der auf eben erwähnten Kenngrößen univariater Statistik herausgearbeitet werden können. Zudem sind Formeln zur Berechnung zwei- oder mehrdimensionaler Zusammenhangsmaße sehr oft deutlich leichter darzustellen, wenn auf eindimensionale Kenngrößen zurückgegriffen werden kann.

Auswertung eindimensionaler Merkmale (univariate Analyse)

Lageparameter

- Modalwert (= Modus)
- arithmetisches Mittel
- geometrisches Mittel
- harmonisches Mittel
- Median
- Quartile und Quantile

Streuparameter

- Spannweite und Quartilsabstand
- empirische Varianz und Standard-abweichung
- mittlere absolute Abweichung (vom Median)
- Variationskoeffizient

Abb. 2-4: Typische Kenngrößen der univariaten Datenanalyse

Abgesehen von Praktikabilitäts- bzw. Effizienzgründen hat die Messung *mehrerer* Merkmale in einem Durchgang ihren Vorteil jedoch hauptsächlich darin, dass durch eine entsprechende Analyse Zusammenhänge zwischen den Größen – sofern vorhanden – sichtbar gemacht werden können. Dazu müssen folglich mehrere Merkmale *gleichzeitig* in die statistische Untersuchung miteinfließen. Diese mehrdimensionale Statistik erfordert jedoch „z. T. recht anspruchsvolle mathematische Methoden, die nur noch auf einem Rechner ausgeführt werden können. Daher beschränken wir uns [...] auf die [...] simultane Auswertung von zwei Merkmalen. Hierbei sind die Ergebnisse und Rechenverfahren [...] gut nachvollziehbar.“ (SIBBERTSEN & LEHNE 2015, S. 104).

Wird ein Zusammenhang zwischen zwei Merkmalen vermutet, stellt sich natürlich vorrangig die Frage nach seiner tatsächlichen Existenz, der Richtung (Positiv-/ Negativtrend), der Art (linear, quadratisch/polynomial, logarithmisch, logistisch, ...) sowie der Stärke des Zusammenhangs.

2.2.2 *Zweidimensionale Häufigkeitstabellen: Kontingenz- bzw. Korrelationstabellen*

Im Folgenden seien X und Y zwei Merkmale mit den jeweiligen Ausprägungen x_j (j = 1, ..., m) und y_k (k = 1, ..., r). Da jede Merkmalsausprägung x_j als Kombination mit y_k auftreten kann, gibt es folglich $m \cdot r$ Wertepaare (x_j, y_k). Werden nun an n Merkmalsträgern Daten in Form eines solchen Wertepaares erhoben, stehen für die Analyse aus diesem Pool aller möglichen Wertepaare die Paare (x_i, y_i) für $i \in \{1, ..., n\}$ zur Verfügung. Selbstverständlich können unter diesen n Wertepaaren bestimmte Paare mehrfach auftreten, so dass die Anzahl dieser Wertepaare als absolute Häufigkeiten in einer Tabelle dargestellt werden können. So eine **zweidimensionale** (bivariate) Kreuz- bzw. **Häufigkeitstabelle** enthält im Vergleich zur univariaten Häufigkeitstabelle die Häufigkeiten von Ausprägungs*paaren*. Sie gibt einen guten „Überblick darüber, wie die einzelnen Ausprägungen im Datensatz verteilt sind.“ (CRAMER & KAMPS, S. 17). Die Abbildung 2-5 zeigt eine solche Häufigkeitstabelle in allgemeiner Form.

	y_1	y_2		y_r	$\sum$
x_1	$H(x_1, y_1)$	$H(x_1, y_2)$		$H(x_1, y_r)$	$H(x_1)$
x_2	$H(x_2, y_1)$	$H(x_2, y_2)$		$H(x_2, y_r)$	$H(x_2)$
x_m	$H(x_m, y_1)$	$H(x_m, y_2)$		$H(x_m, y_r)$	$H(x_m)$
$\sum$	$H(y_1)$	$H(y_2)$		$H(y_r)$	n

Abb. 2-5: Allgemeine Form einer bivariaten Häufigkeitstabelle inkl. Randhäufigkeiten. $H(x_j, y_k)$ gibt jeweils die absolute Häufigkeit einer auftretenden Merkmalskombination an.

Nicht selten werden in einer erweiterten Form einer solchen Kreuz- bzw. Häufigkeitstabelle in der rechten Spalte bzw. in der untersten Zeile die so genannten Randhäufigkeiten mit angegeben: $H(x_i)$ am rechten Tabellenrand gibt z. B. die auftretende Anzahl des Merkmals X in seiner Ausprägung x_i an. In der obigen Abbildung 2-5 sind die Randhäufigkeiten zur besseren Übersicht blau hinterlegt. In der äußersten Zelle rechts unten zeigt die Zahl n die Summe aller horizontalen bzw.

vertikalen Werte an: $n = \sum_{k=1}^{r} H(y_k) = \sum_{j=1}^{m} H(x_j)$. Sie stellt den Erhebungs- oder Stichprobenumfang dar.

„Da ja absolute Häufigkeiten für sich allein schlecht zu interpretieren sind“ (SIBBERTSEN & LEHNE 2015, S. 107), werden in den meisten Fällen die relativen Häufigkeiten abgedruckt. Dividiert man jeweils die absolute Häufigkeit $H(x_j, y_k)$ einer auftretenden Merkmalskombination durch die Anzahl n aller Merkmalsträger (= Erhebungsumfang), ergibt sich der entsprechende relative Häufigkeitswert $h(x_i; y_i)$. Es gilt also $h(x_i, y_i) = \frac{H(x_i, y_i)}{n}$ und für die Randhäufigkeiten ebenso $h(x_i) = \frac{H(x_i)}{n}$ und $h(y_i) = \frac{H(y_i)}{n}$. Die Abbildung 2-6 zeigt das Pendant zur Tabelle der Abbildung 2-5 mit relativen anstatt absoluten Häufigkeitswerten.

	y_1	y_2		y_r	Σ
x_1	$h(x_1, y_1)$	$h(x_1, y_2)$		$h(x_1, y_r)$	$h(x_1)$
x_2	$h(x_2, y_1)$	$h(x_2, y_2)$		$h(x_2, y_r)$	$h(x_2)$
x_m	$h(x_m, y_1)$	$h(x_m, y_2)$		$h(x_m, y_r)$	$h(x_m)$
Σ	$h(y_1)$	$h(y_2)$		$h(y_r)$	1 (bzw. 100 %)

Abb. 2-6: Beispiel für eine bivariate Häufigkeitstabelle mit Angabe der relativen Häufigkeiten

Die zwei Tabellen in Abbildung 2-7 (S. 19) verdeutlichen die beiden Darstellungsmöglichkeiten an einem konkreten Beispiel. Dazu wurde das Ergebnis einer Befragung von 70 Sechstklässlern zu ihrer Schulnote im Fach Mathematik und Englisch auf beide oben beschriebenen Arten dargestellt: Die linke Tabelle enthält die absoluten Häufigkeiten des jeweiligen Ausprägungspaars, die rechte die relativen Häufigkeiten.

		Englischnote					
		1	**2**	**3**	**4**	**5**	
Mathematiknote	**1**	1	2	3	–	–	6
	2	3	5	8	2	1	19
	3	2	4	12	9	2	29
	4	2	2	2	6	1	13
	5	–	–	2	1	–	3
		8	13	27	18	4	70

		Englischnote					
		1	**2**	**3**	**4**	**5**	
Mathematiknote	**1**	1,4 %	2,9 %	4,3 %	0,0 %	0,0 %	8,6 %
	2	4,3 %	7,1 %	11,4 %	2,9 %	1,4 %	27,1 %
	3	2,9 %	5,7 %	17,1 %	12,9 %	2,9 %	41,4 %
	4	2,9 %	2,9 %	2,9 %	8,6 %	1,4 %	18,6 %
	5	0,0 %	0,0 %	2,9 %	1,4 %	0,0 %	4,3 %
		11,4 %	18,6 %	38,6 %	25,7 %	5,7 %	100 %

Abb. 2-7: Zwei Beispiele für bivariate Häufigkeitstabellen

Üblicherweise werden zweidimensionale Tabellen oder Tafeln dieser Art **Korrelationstabellen** genannt, außer mindestens eines der Merkmale ist nominal skaliert, dann spricht man von einer **Kontingenztabelle** oder -tafel. Die Abbildung 2-8 zeigt als Beispiel eine Kontingenztabelle mit den absoluten Häufigkeiten der Ausprägungspaare (Geschlecht | Schulnote).

	1	**2**	**3**	**4**	**5**	**6**	
m	0	2	5	6	3	1	17
w	2	3	5	2	2	0	14
	2	5	10	8	5	1	31

Abb. 2-8: Beispiel für eine Kontingenztabelle

2.2.3 Der Kontingenzkoeffizient

Auch wenn sich aus einer Kontingenz- bzw. Korrelationstabelle durchaus erste Anhaltspunkte über die Stärke eines Zusammenhangs zweier Variablen herauslesen lassen (vgl. CLEFF 2014, S. 75), gestaltet sich die Einschätzung umso schwieriger, je mehr Zeilen und Spalten die Tabelle aufweist. Ein Parameter bzw. eine Kennzahl, die die Stärke des Zusammenhangs zum Ausdruck bringt, erscheint an dieser Stelle wünschenswert. Eine in der Statistik gängige Maßzahl für die Angabe der Zusammenhangsstärke ist der so genannte **Kontingenzkoeffizient** (nach Pearson). Er wird verwendet, „wenn mindestens eines der beiden Merkmale nominales Skalenniveau aufweist.“ (SIBBERTSEN & LEHNE 2015, S. 116). Generell ist zu beachten, dass bei der Analyse von Merkmalspaaren stets dasjenige Verfahren Anwendung findet, das für das Merkmal mit dem geringeren Skalenniveau ausgerichtet ist (vgl. Übersicht der Abb. 2-14 auf S. 26). Selbstverständlich kann man auch bei metrisch skalierten Daten den Kontingenzkoeffizienten berechnen, allerdings werden „dann Informationen über die Abstände verschenkt“. (SIBBERTSEN & LEHNE 2015, S. 138).

In der Regel wird in der Statistik von einem Parameter, der ein Zusammenhangsmaß angibt, erwartet, dass der errechnete Wert (bzw. der Betrag davon) in dem Intervall [0; 1] liegt: 0 bedeutet, es liegt kein Zusammenhang zwischen den beiden Merkmalen vor; 1 bedeutet, es liegt ein eindeutiger Zusammenhang vor. Der bereits erwähnte Kontingenzkoeffizient liefert zwar in seiner ursprünglichen Definition Werte zwischen 0 und 1, schöpft aber das Intervall [0; 1] nicht vollständig aus, da er von der Dimension der betrachteten Kontingenztabelle abhängt (vgl. CRAMER & KAMPS 2014, S. 104). Dies erschwert eine eindeutige Interpretation und somit auch den Vergleich zweier Kontingenztafeln unterschiedlicher Dimension. Darum wird bei der Ermittlung eines Zusammenhangsmaßes (bei mindestens einem nominalen Merkmal) häufig der ***korrigierte* Kontingenzkoeffizient** verwendet, der meist mit C^* (manchmal auch mit K^* oder C_{korr}) angegeben wird und den eben beschriebenen Anforderungen genügt ($C^* \in [0; 1]$). Die Formeln für die beiden Kontingenzkoeffizienten, angepasst an die bisherigen Schreibweisen und Indizes, können der Abbildung Abb. 2-9 entnommen werden. Weitere Erläuterungen (z. B. zur Berechnung und Bedeutung des Chi-Quadrat-Koeffizienten als Vorbereitung) und Berechnungsbeispiele zur Ermittlung dieser typischen Kenngrößen finden sich z. B. in CLEFF (2014) oder SIBBERTSEN und LEHNE (2015).

Formel zur Berechnung des Kontingenzkoeffizienten C:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Korrigierter Kontingenzkoeffizient:

$$C^* = \sqrt{\frac{\chi^2}{\chi^2 + n} \cdot \frac{\min(m, r)}{\min(m, r) - 1}}$$

Abb. 2-9: Formeln zur Berechnung des Kontingenzkoeffizienten C sowie des korrigierten Kontingenzkoeffizienten C*

Mit Hilfe des (korrigierten) Kontingenzkoeffizienten ist es also möglich, den Zusammenhang und dessen Intensität zwischen zwei Variablen durch einen Parameter bzw. Wert aus dem Intervall [0; 1] anzugeben, jedoch lassen sich dabei weder Aussagen über die Form der Abhängigkeit (linear/nicht linear) noch über deren Wirkungsrichtung (positiv/negativ) treffen. Abhilfe kann hier teilweise der Korrelationskoeffizient schaffen; dieser setzt allerdings bei *beiden* Merkmalen ein metrisches Skalenniveau voraus.

2.2.4 Das Streudiagramm

Sind beiden Merkmale X und Y des Paars (X, Y) metrisch skaliert, bietet es sich noch *vor* der Berechnung entsprechender Zusammenhangsmaße für dieses Messniveau an, die Daten zu visualisieren, um einen ersten Eindruck über die Wirkungsrichtung, die Form und die Stärke eines möglichen Zusammenhangs zu gewinnen. Die einfachste und wohl bekannteste Darstellungsform dieser Art ist das **Streudiagramm** (engl.: *scatterplot*) bzw. Dabei wird für $i \in \{1, ..., n\}$ jedes Beobachtungspaar (x_i, y_i) als Punkt in ein zweidimensionales Koordinatensystem eingetragen. Der so entstandenen Veranschaulichung von n Punkten in einer Ebene (sie wird deswegen auch Punktwolke genannt) kann man oftmals per Augenschein entnehmen, ob überhaupt ein Zusammenhang zwischen den beiden Variablen vorliegt. Darüber hinaus kann man Informationen über die drei folgenden Aspekte gewinnen:
(Die folgenden Erläuterungen mit der entsprechenden Abbildung dazu sind größtenteils wortwörtlich aus CLEFF (2014, S. 95 f) entnommen.)

1. „Die **Richtung** des Zusammenhangs: So lassen sich positive, negative und fehlende Trends feststellen“. Korrespondieren wachsende Werte von X mit wachsenden Werten von Y, spricht man von einem positiven Trend. Bei negativen Zusammenhängen sinken die Werte von Y bei einer Zunahme der X-Werte. Lässt sich keine Systematik dieser Art erkennen, lässt dies auf einen fehlenden oder äußerst schwachen Zusammenhang zwischen den Variablen schließen.
2. „Die **Form** des Zusammenhangs […]: Es lassen sich lineare oder nicht lineare Zusammenhänge unterscheiden.“

3. „Die **Stärke** eines Zusammenhangs […]: Liegen die Punkte sehr dicht an einer Geraden, dann ist der lineare Zusammenhang stärker als bei einem Streudiagramm, bei dem die Punkte weit von der Geraden entfernt liegen."

Die Abbildung 2-10 illustriert durch verschiedene Punktwolken diese Aspekte des Zusammenhangs zwischen zwei Merkmalen.

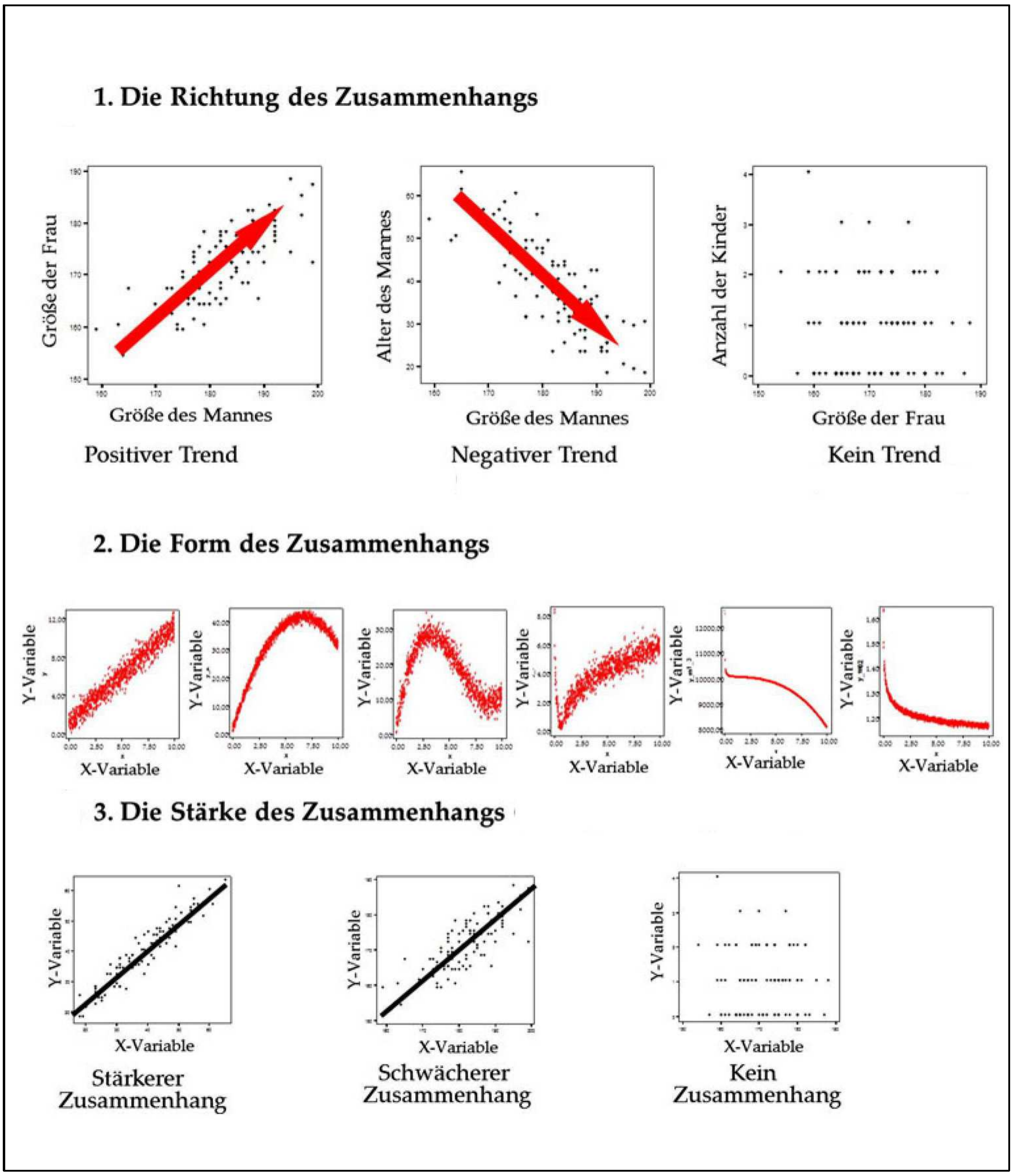


Abb. 2-10: Unterschiedliche Aspekte bei Streudiagrammen (CLEFF 2015, S. 96)

Auch wenn sich in der Regel Streudiagramme relativ einfach mithilfe diverser Software realisieren lassen, ist bei deren Interpretation doch stets Vorsicht geboten: Eine Veränderung der Achsenskalierung kann beispielsweise dazu führen, dass die Stärke des Zusammenhangs falsch eingeschätzt wird. Die Abbildung 2-11 zeigt auf der Grundlage desselben Datensatzes zweimal denselben Sachverhalt, allerdings wird man in der rechten Veranschaulichung einen deutlich stärkeren Zusammenhang vermuten als in der linken, da die Wertepaare scheinbar in deutlich geringerem Ausmaß um die Gerade streuen.

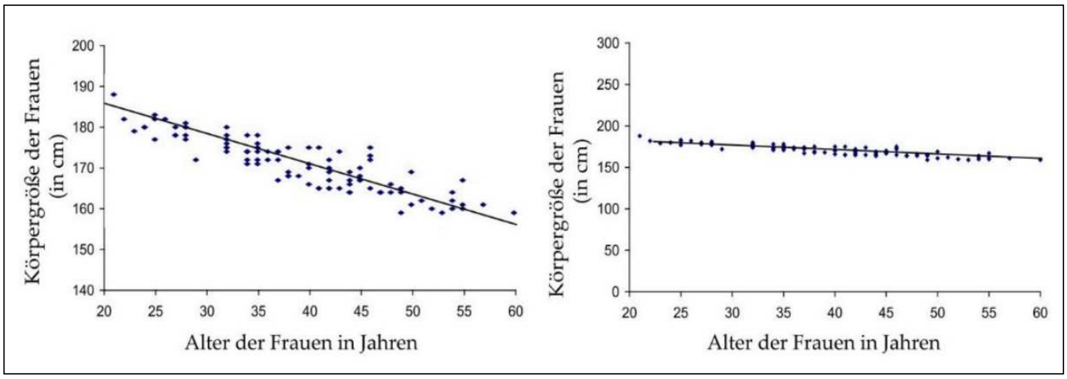


Abb. 2-11: Unterschiedliche Darstellung des gleichen Sachverhalts (nach CLEFF 2015, S. 97)

Abgesehen von einer unter Umständen auch bewusst vorgenommenen Manipulation der Skalierung unterliegt jedes Streudiagramm auch der subjektiven Einschätzung des Betrachters. Die Forderung bzw. Angabe einer Maßzahl, die ein „ungetrübtes Bild über einen Zusammenhang von zwei metrischen Variablen liefert" (CLEFF 2014, S. 97), scheint deshalb sinnvoll.

2.2.5 Der Korrelationskoeffizient und die Rangkorrelation

Mit Hilfe des Korrelationskoeffizienten nach Bravais-Pearson lässt sich zwar keine Aussage über die Form eines Zusammenhangs zweier metrisch skalierter Merkmale treffen, wohl aber über die Richtung und dessen Stärke. „Anders als der Kontingenzkoeffizient basiert er nicht auf der Häufigkeit der Merkmalsausprägungen von (X, Y), sondern direkt auf den Beobachtungswerten" (CRAMER & KAMPS 2014, S. 105). Dadurch gestaltet sich die Berechnung auch bei bereits wenigen Wertepaaren als aufwändig, da für die Ermittlung des Korrelationskoeffizienten nach

Bravais-Pearson (in der Literatur häufig kurz mit r oder r_{XY} bezeichnet) die Kovarianz, also die durchschnittliche Abweichung zum so genannten bivariaten Schwerpunkt, als auch die Standardabweichung von X und Y benötigt werden. Die Formel für diesen Korrelationskoeffizienten in einer vereinfachten Schreibweise kann der Abbildung 2-12 entnommen werden; für weitere Erläuterungen (z. B. zur Kovarianz S_{XY}) und Berechnungsbeispiele sei auch an dieser Stelle auf die bereits genannte Literatur verwiesen (CLEFF 2014 oder SIBBERTSEN & LEHNE 2015).

Formel zur Berechnung des Korrelationskoeffizienten r nach Bravais-Pearson:

$$r = \frac{s_{XY}}{s_X s_Y} \qquad \text{mit} \qquad s_{XY} = \frac{1}{n} \sum_{i=0}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

$$\text{und} \qquad s_X = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (x_i - \bar{x})^2} \qquad \text{bzw.} \qquad s_Y = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (y_i - \bar{y})^2}$$

Abb. 2-12: Formel zur Berechnung des Korrelationskoeffizienten nach Bravais-Person

Der Korrelationskoeffizient nach Bravais-Pearson liefert als Ergebnis einen Wert aus dem Intervall [–1; 1]. An den beiden Intervallgrenzen lässt sich sein Verhalten folgendermaßen charakterisieren (vgl. CRAMER & KAMPS 2014, S. 110 f):

- $r = 1$, wenn die Beobachtungswerte auf einer Geraden mit positiver Steigung liegen,
- $r = -1$, wenn ein negativer Zusammenhang besteht, d. h. die Beobachtungswerte liegen auf einer fallenden Geraden.

Für $r = 0$ besteht zwischen den untersuchten Merkmalen zumindest kein *linearer* Zusammenhang. Die Abbildungen veranschaulichen jeweils den Idealfall $r = \pm 1$ und zweimal den Fall $r \approx 0$:

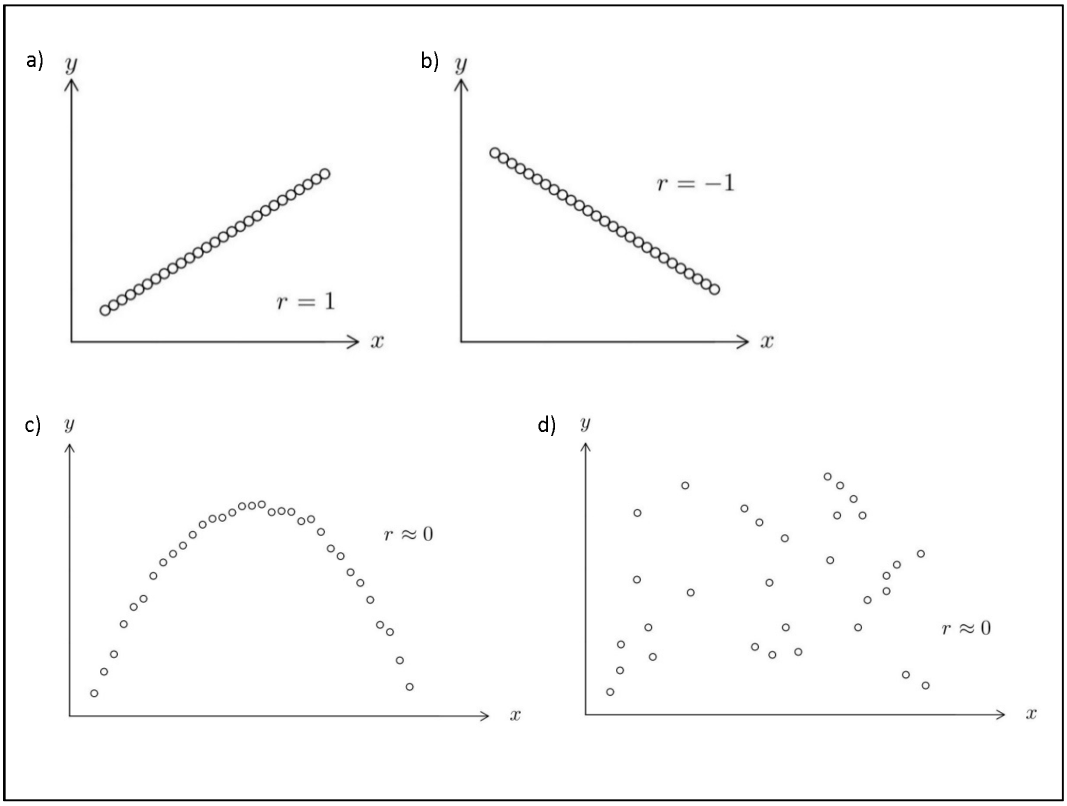


Abb. 2-13: Exakter positiver und negativer Zusammenhang sowie zwei Beispiele für r ≈ 0. (Darstellungen a, b und c nach SIBBERTSEN & LEHNE 2015, S. 130 f)

Bei der Abbildung 2-13 c) fällt ein quadratischer Zusammenhang ins Auge, der in Anbetracht der Tatsache, dass der Korrelationskoeffizient mit r ≈ 0 angegeben ist, verwunderlich erscheint. Dieses Beispiel soll jedoch verdeutlichen, dass der Korrelationskoeffizient nach Bravais-Pearson nur die Stärke *linearer* Zusammenhänge misst. Ein Wert nahe der Null bedeutet also nicht zwangsläufig, „dass überhaupt kein Zusammenhang zwischen den beiden Merkmalen existiert. Es bedeutet lediglich, dass anhand des Datenmaterials kein *linearer* Zusammenhang nachgewiesen werden kann.“ (CRAMER & KAMPS 2014, S. 112).

In den bisherigen Ausführungen bleibt bisher die Fragen offen, welche Mittel zur Verfügung stehen, wenn ein Zusammenhangsmaß für bivariate Merkmale (X, Y) berechnet werden soll, bei dem X und Y mindestens ordinales Messniveau aufweisen (vgl. CRAMER & KAMPS 2014, S. 113). „Oder was wäre zu tun, wenn der Zusammenhang nicht linear, sondern nur monoton ist?“ (CLEFF 2014, S. 101). In diesem Fall wird in der Regel der so genannte **Rangkorrelationskoeffizient** nach Spearman (übliche Schreibweisen: r_S, r_{Sp} oder ρ) verwendet. Ähnlich wie der Korrelationskoeffizient nach Pearson nimmt er ebenfalls Werte aus dem Intervall [−1; 1] an: Für $r_S > 0$ besteht ein monoton wachsender Zusammenhang, für $r_S < 0$ ein monoton fallender Zusammenhang und für $r_S = 0$ besteht kein monotoner Zusammenhang, das heißt, je „mehr sich der Wert des Koeffizienten null nähert, umso mehr weichen die Wertepaare von einem […] monotonen Zusammenhang ab." (CRAMER & KAMPS 2014, S. 104). Im Idealfall $r_S = 1$ beispielsweise würden „alle Beobachtungswerte […] auf einer in ihrer Steigung variierenden aber stets ansteigenden Kurve" liegen (CRAMER & KAMPS 2014, S. 102).

Die Formel zur Berechnung der Rangkorrelationskoeffizienten ist der des Korrelationskoeffizienten nach Bravais-Pearson sehr ähnlich, allerdings treten – da man bei ordinalen Daten die Abstände ja nicht bestimmen kann – an die Stelle der ursprünglichen Beobachtungswerte die entsprechenden Ränge, d. h. „x_i wird durch $rg(x_i)$ (= Rang von x_i) und y_i wird durch $rg(y_i)$ (= Rang von y_i) ausgetauscht." (SIBBERTSEN & LEHNE 2015, S. 134). Abgesehen davon, dass r_S nicht die Annahme benötigt, dass der Zusammenhang zwischen zwei Merkmalen linear ist, stellt er bei Ausreißerpaaren das deutlich robustere Zusammenhangsmaß dar (vgl. CLEFF 2014, S. 102).

Die Abbildung 2-14 gibt einen tabellarischen Überblick, welches Zusammenhangsmaß je nach vorliegendem Skalenniveau der einzelnen Merkmale angewendet werden kann. Dabei gilt: „Die Beobachtungsgröße mit dem geringsten Messniveau bestimmt [...] die anzuwendende Methode." (CRAMER & KAMPS 2014, S. 91).

		Skalenniveau von Y		
		nominal	ordinal	kardinal
Skalenniveau von X	nominal	C*	C*	C*
	ordinal	C*	r_S	r_S
	kardinal	C*	r_S	r

C*: korrigierter Kontingenzkoeffizient,
r_S: Rangkorrelationskoeffizient nach Spearman
r: Korrelationskoeffizient nach Bravais-Pearson

Abb. 2-14: Zusammenhangsmaße in Abhängigkeit vom Skalenniveau (nach SIBBERTSEN & LEHNE 2015, S. 138)

„Zu beachten ist weiter, dass die drei Zusammenhangsmaße unterschiedliche Formen der Abhängigkeit messen“ (SIBBERTSEN & LEHNE 2015, S. 138):

- Bravais-Pearson-Korrelationskoeffizient r: linearer Zusammenhang
- Rangkorrelationskoeffizient nach Spearman r_S: monotoner Zusammenhang
- korrigierter Kontingenzkoeffizient C*: allgemeiner Zusammenhang

„Das Zusammenhangsmaß wird man daher auch danach auswählen, welche Art der Abhängigkeit vermutet wird. Geht man von einem streng-monotonen Zusammenhang aus, bietet sich auch bei metrischen Daten der Rangkorrelationskoeffizient an.“ (SIBBERTSEN & LEHNE 2015, S. 138).

Abschließend sei vermerkt, dass eine Einteilung, ab welchem Wert ein Koeffizient einen *hohen* oder *starken* Zusammenhang darstellt, mit Vorsicht zu genießen ist. In einschlägiger Fachliteratur ist zudem je nach Autor oft eine andersgelagerte Einteilung der Wertebereiche für die Stärke des Zusammenhangs zu finden.

2.2.6 Kausalität und Scheinkorrelation

Auch wenn der Betrag eines Korrelationskoeffizienten nahe 1 einen starken (linearen) Zusammenhang zwischen zwei Merkmalen belegt, so gibt er doch keine Auskunft darüber, welches Merkmal das jeweils andere beeinflusst. Ist die Variable X der *Verursacher* (**Regressor**) und wirkt auf die (dann abhängige) Variable Y (**Regressand**) oder umgekehrt? Oder beeinflussen sich die beiden Merkmale unter Umständen sogar gegenseitig? Die Frage der **Kausalität** lässt sich nur im jeweiligen Sachkontext (vgl. EICHLER & VOGEL 2013, S. 122), manchmal auch gar nicht beantworten. Bei verheirateten Paaren z. B. besteht nachweislich eine starke Korrelation zwischen dem Alter des Bräutigams und dem Alter der Braut. Bestimmt nun das Alter des Bräutigams das Alter der Braut oder umgekehrt (vgl. CLEFF 2014, S. 135)?

Wenn eine Ursache-Wirkungsbeziehung auf den ersten Blick unsinnig erscheint und sich die Kausalität nicht erschließen lässt, kann hinter einem hohen Zusammenhangsmaß auch nur der Zufall stecken. Es kann auch sein, dass eine dritte, nicht berücksichtigte Variable die Korrelation zu vertreten hat, indem sie *beide* korrelierenden Merkmale gleichzeitig beeinflusst. Oder die tatsächlich kausale

Variable wirkt bereits im Vorfeld bzw. indirekt auf eines der beiden untersuchten Merkmale ein. Ein Beispiel hierzu:
Es lässt sich eine positive Korrelation zwischen Körpergröße und Alkoholkonsum nachweisen. Die Schlussfolgerung, alle großen Menschen seien Trinker, scheint jedoch abwegig. Dieser scheinbare Zusammenhang kommt jedoch dadurch zustande, dass „die Häufigkeit des Alkoholkonsums eindeutig auch vom Geschlecht abhängt: Männer weisen einen höheren Alkoholkonsum auf." (CLEFF 2014, S. 125). Da Männer in der Regel größer sind als Frauen, ist die eigentlich kausale Variable eher das Geschlecht und nicht die Körpergröße (vgl. CLEFF 2014, S. 125 f).
Einem errechneten Zusammenhangsmaß kann letztendlich auch gar keine tatsächliche Abhängigkeit zu Grunde liegen (**Scheinkorrelation**); es ist durchaus möglich, dass sich aus einem Sachzusammenhang kein kausaler Zusammenhang ableiten lässt. „Ein Beispiel dafür wäre ein hoher Kontingenzkoeffizient zwischen Augenfarbe und Schulabschluss." (SIBBERTSEN & LEHNE 2015, S. 139).

2.2.7 Lineare Regression

Häufig ist man nicht nur an einer Maßzahl zur Bestimmung eines Zusammenhangs interessiert, sondern an einer mathematischen Funktion, die diesen Zusammenhang möglichst exakt beschreibt. In den seltensten Fällen lassen sich Zusammenhänge eindeutig durch eine (nicht zwangsläufig lineare) Funktion beschreiben, häufig aber zumindest annähernd. Der Versuch, die Abhängigkeit quantitativer (= metrischer) Merkmale als *Regel* zu beschreiben, wird als **Regression** bezeichnet. Entsprechend nennt man eine Funktion, die die Tendenz eines Zusammenhangs möglichst exakt zu beschreiben versucht, **Regressionsfunktion**.
Da das Einfügen hypothetischer Werte in ein bestehendes Datengefüge auf dem Grundgedanken der Regression beruht und dies ein Hauptuntersuchungsfeld dieser Arbeit darstellt, wird diese Thematik an dieser Stelle ausführlicher betrachtet.
Nimmt der Bravais-Pearson-Korrelationskoeffizient beispielsweise den Wert −0,93 an, so ist dadurch zwar bereits die Richtung (negativer Trend), die Art des Zusammenhangs (linear) als auch dessen Stärke (stark korrelierend) bekannt, allerdings ist es häufig erstrebenswert, eine Prognose für weitere Merkmalspaare anstellen zu können: Welchen Wert nimmt die entsprechende Merkmalsausrägung y_i eines

Wertepaares (x_i, y_i) bei einem nicht beobachteten Wert x_i an (i = 1, ..., n)? Genau solche Vorhersagen sollen durch eine entsprechende Modellfunktion ermöglich werden. Sofern ein Zusammenhang zwischen den beobachteten Merkmalspaaren überhaupt durch eine Funktion erfasst werden kann, liegen die prognostizierten Merkmalspaare (X, Y) im Idealfall direkt auf dieser Regressionsfunktion, in den meisten Fällen allerdings zumindest *in der Nähe* davon.

Bereits vor der Bestimmung einer Regressionsfunktion sollte auch die Kausalität bzw. die Wirkrichtung eines Zusammenhangs hinterfragt werden: Führt „eine Veränderung des Merkmals X zu einer Veränderung von Y [...] oder [...] andersherum“? (SIBBERTSEN & LEHNE 2014, S. 141). Es sei vor „dem Schluss gewarnt, die Regression beweise die Kausalität von Zusammenhängen: Die Kausalität [...] muss zuvor theoretisch abgeleitet werden, bevor sie empirisch mit Hilfe der Regression bewiesen werden kann.“ (CLEFF 2014, S. 135; vgl. Kap. 2.2.6).

Anhaltspunkte für eine bestimmte Abhängigkeitsstruktur zweier metrischer Merkmale X und Y ergeben sich in der Regel aus theoretischen Überlegungen oder durch Auswertung des Streudiagramms (vgl. CRAMER & KAMPS 2014, S. 117). In der Praxis wird zwei metrischen Merkmalen, so wie hier in den nachfolgenden Ausführungen auch, anfangs oft ein linearer Zusammenhang unterstellt. Das scheint nicht nur aufgrund praktischer Erfahrungen plausibel, sondern auch aufgrund der Tatsache, dass man Zusammenhänge oft nur in beschränkten Intervallen untersucht. Abgesehen davon, dass sich die Berechnung einer Regressionsgerade bei linearem Ansatz noch relativ leicht nachvollziehen lässt (vgl. CRAMER & KAMPS 2014, S. 120), spricht für eine erste Annahme einer linearen Funktion auch das von EICHLER und VOGEL zitierte „wissenschaftliche Sparsamkeitsprinzip *Ockhams Rasier messer* [...]: Wenn das einfache Modell ausreicht, dann erhält es auch den Vorzug.“ (EICHLER & VOGEL 2013, S. 121). „Für die Datenarbeit mit parametrischen Funktionen spezifiziert Erickson [...] dieses Prinzip folgendermaßen: *»Make your models with as many parameters as you need – but no more.«*“ (EICHLER & VOGEL 2013, S. 138 f).

Für die folgenden Überlegungen sei also ein lineares Regressionsmodell angenommen, bei dem ein metrisch skaliertes Merkmal X linear auf ein ebenfalls metrisch skaliertes Merkmal Y einwirkt. Wird nun versucht, die abhängige Variable Y

(Regressand) in Abhängigkeit des erklärenden bzw. unabhängigen Merkmals X (Regressor) über die Gleichung $Y = f(X)$ darzustellen, wird das vermutlich nur annähernd gelingen, da die Regressionswerte $\hat{y}_i = f(x_i)$ für $i \in \{1, \ldots, n\}$ im Allgemeinen um die Regressionsgerade streuen; dies kann Messfehlern, aber auch natürlichen Schwankungen in den Eigenschaften der Merkmalsträger geschuldet sein (vgl. CRAMER & KAMPS 2014, S. 118). Um der auftretenden Abweichung der Schätzwerte $\hat{y}_i$ an der Stelle x_i Rechnung zu tragen, wird in die Funktionsgleichung die additive Komponente u_i miteinbezogen. Diese Differenz aus $\hat{y}_i$ und y_i ($i = 1, \ldots, n$) nennt man **Residuum**. Für jedes Datenpaar (x_i, y_i) ergibt sich dann die Beziehung $y_i = f(x_i) + u_i$ für $i \in \{1, \ldots, n\}$. Jeder Punkt kann also „als Kombination des Ergebnisses der Regressionsgeraden $\hat{y}$ und des jeweiligen Residuums ausgedrückt werden“ (CLEFF 2014, S. 139).

Ziel einer gelungenen Approximation ist es nun, die Summe aller Residuen u_i so zu minimieren, so dass die Regressionsgerade *„so nah wie möglich* an möglichst vielen Messpaaren *vorbeikommt“* (CLEFF 2014, S. 139). Eine der populärsten Methoden zur Generierung einer solchen (eindeutigen) Ausgleichsgeraden ist die Methode der Kleinsten Quadrate nach C. F. Gauß: Dabei wird versucht, die Summe der quadrierten Abstände zwischen den tatsächlichen y_i und den entsprechenden Punkten auf der Geraden $\hat{y}$ zu minimieren:

$$\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \rightarrow \min$$

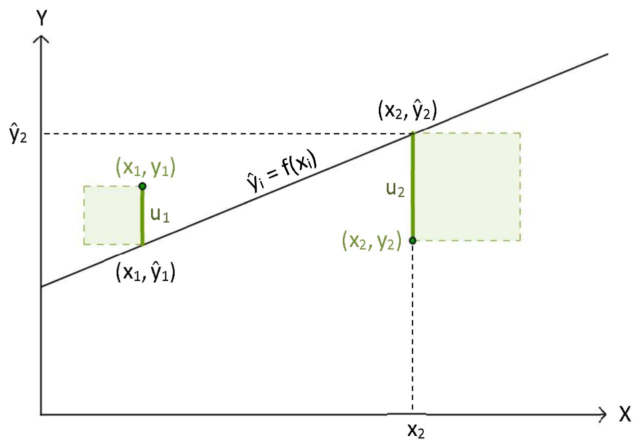


Abb. 2-15: Veranschaulichung der Methode der Kleinsten Quadrate

In Abbildung 2-15 ist diese Methode an zwei Wertepaaren exemplarisch veranschaulicht.

Legt man gemäß der bisherigen Annahme als Regressionsfunktion eine lineare Funktion f(a, b) der Form $\hat{y} = a + bx$ zugrunde, ergibt sich für $\hat{y}_i = a + bx_i$:

$$f(a, b) = \sum_{i=1}^{n} \left(a + bx_i - y_i\right)^2 \rightarrow \min$$

Welche Werte müssen also a und b nun annehmen, so dass die Summe der quadrierten Abweichungen ihr Minimum annimmt? Die typische Vorgehensweise ist die partielle Differentiation nach den beiden Variablen a und b und das anschließende Nullsetzen beider Ableitungen. Das so entstandene lineare Gleichungssystem mit zwei Unbekannten kann dann z. B. durch das Einsetzverfahren gelöst werden. Im vorliegenden Fall scheint das Procedere leichter nachzuvollziehen, wenn man zuerst die erste Ableitung $\frac{\partial f(a,b)}{\partial a}$ nullsetzt, dann umformt und die so entstandene (nach a aufgelöste) Gleichung (I) anschließend gleich in f(a, b) einsetzt, noch bevor man nach b differenziert:

(Der besseren Lesbarkeit wegen wird bis zum Abschluss dieses Kapitels auf die Notation des Summationsindex i = 1, …, n verzichtet.)

$$\frac{\partial f(a,b)}{\partial a} = \sum 2 \cdot \left(a + bx_i - y_i \right)$$

$$\begin{aligned} \sum 2 \cdot \left(a + bx_i - y_i \right) \overset{!}{=} 0 \quad &\Leftrightarrow \quad \sum a + b \sum x_i - \sum y_i = 0 \\ &\Leftrightarrow \quad \sum a + b \cdot n \cdot \tfrac{1}{n} \sum x_i - n \cdot \tfrac{1}{n} \sum y_i = 0 \\ &\Leftrightarrow \quad n \cdot a + b \cdot n \cdot \overline{x} - n \cdot \overline{y} = 0 \\ &\Leftrightarrow \quad a + b\overline{x} - \overline{y} = 0 \\ &\Leftrightarrow \quad a = \overline{y} - b\overline{x} \quad \text{(I)} \end{aligned}$$

$$\begin{aligned} \sum\left(a + bx_i - y_i\right)^2 &= \sum\left(\overline{y} - b\overline{x} + bx_i - y_i\right)^2 = \sum\left((x_i - \overline{x})b - (y_i - \overline{y})\right)^2 \\ &= \sum\left((x_i - \overline{x})^2 b^2 - 2(x_i - \overline{x})b\left(y_i - \overline{y}\right) + (y_i - \overline{y})^2\right)^2 \\ &= \sum(x_i - \overline{x})^2 b^2 - 2\sum(x_i - \overline{x})\left(y_i - \overline{y}\right)b + \sum(y_i - \overline{y})^2 \end{aligned}$$

$$\Rightarrow \frac{\partial f(a,b)}{\partial b} = 2\sum(x_i - \overline{x})^2 b - 2\sum(x_i - \overline{x})\left(y_i - \overline{y}\right)$$

$$\begin{aligned} 2\sum(x_i - \overline{x})^2 b - 2\sum(x_i - \overline{x})\left(y_i - \overline{y}\right) \overset{!}{=} 0 &\Leftrightarrow 2\sum(x_i - \overline{x})^2 b = 2\sum(x_i - \overline{x})\left(y_i - \overline{y}\right) \\ &\Leftrightarrow b = \frac{\sum(x_i - \overline{x})\left(y_i - \overline{y}\right)}{\sum(x_i - \overline{x})^2} \end{aligned}$$

Da man $\sum(x_i - \bar{x})(y_i - \bar{y})$ durch ein paar weitere Umformungen (und der Tatsache, dass z. B. $\sum x_i = n \cdot \bar{x}$) als $\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}$ darstellen kann, bestimmt man die Koeffizienten der Gleichung $\hat{y}_i = a + bx_i$ also mit:

$$b = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum(x_i - \bar{x})^2} \quad \text{und} \quad a = \bar{y} - b\bar{x}$$

Dazu eine Beispielaufgabe:

Die Abbildung 2-16 zeigt das Ergebnis einer Umfrage nach der Höhe des wöchentlichen Taschengelds in Bezug auf das Alter von Jugendlichen. Gesucht ist die Funktionsgleichung einer linearen Regressionsgeraden, die den Zusammenhang zwischen den beiden Größen beschreibt.

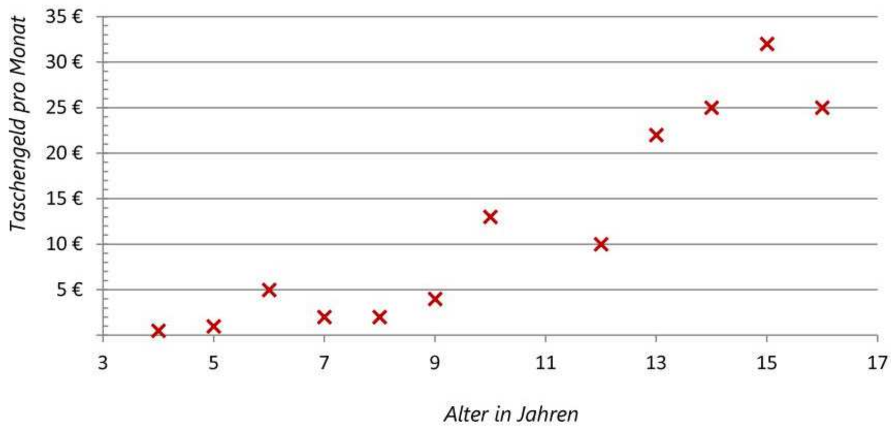


Abb. 2-16: Streudiagramm mit den Ergebnissen einer kleinen Umfrage nach der Höhe des Taschengelds

Für die Berechnung werden – wie in Abbildung 2-17 (S. 33) dargestellt – die abgelesenen Wertepaare in eine Tabelle übertragen und die notwendigen Teilergebnisse ($\bar{x}$, $\bar{y}$, $\sum x_i y_i$ und $\sum(x_i - \bar{x})^2$) berechnet. Für die Regressionsfunktion ergeben sich somit die Koeffizienten

$$b \approx \frac{1869 - 12 \cdot 9{,}9167 \cdot 11{,}7917}{180{,}1967} \approx \frac{465{,}7917}{180{,}1967} \approx 2{,}5746 \quad \text{und}$$

$$a \approx 11{,}7917 - 2{,}5746 \cdot 9{,}9167 \approx -13{,}7400$$

Die Funktionsgleichung lautet also $\hat{y} \approx 2{,}57x - 13{,}74$.

Die Abbildung 2-18 zeigt das um diese Regressionsgerade ergänzte Streudiagramm aus Abbildung 2-16 (S. 32).

Alter in Jahren (x)	Taschengeld in € (y)	$x_i \cdot y_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
4	0,50	2	-5,9167	35,0069
5	1,00	5	-4,9167	24,1736
6	5,00	30	-3,9167	15,3403
7	2,00	14	-2,9167	8,5069
8	2,00	16	-1,9167	3,6736
9	4,00	36	-0,9167	0,8403
10	13,00	130	0,0833	0,0069
12	10,00	120	2,0833	4,3403
13	22,00	286	3,0833	9,5069
14	25,00	350	4,0833	16,6736
15	32,00	480	5,0833	25,8403
16	22,00	400	6,0833	37,0069
$\bar{x} \approx 9{,}9167$	$\bar{y} \approx 11{,}7917$	Σ: 1869		Σ: 180,1967

Abb. 2-17: Berechnung der notwendigen Teilergebnisse zur Ermittlung der Parameter a und b der Regressionsgeraden $\hat{y} = a + bx$

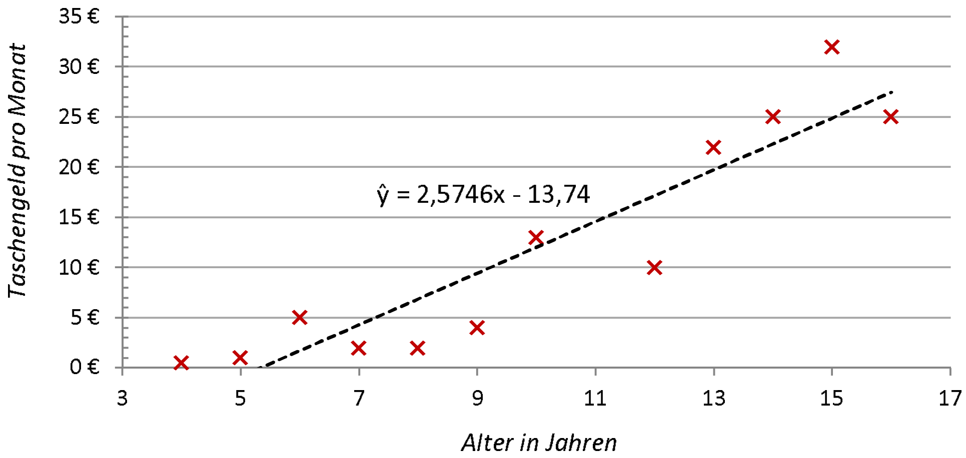


Abb. 2-18: Darstellung der (hier von Excel® automatisch ermittelten) Regressionsgeraden für die Beispielaufgabe „Taschengeld“

2.2.8 Residuenanalyse und nicht-lineare Regression

Zu Recht darf nach der Schilderung der rechnerischen Ermittlung der Ausgleichsgeraden die Frage gestellt werden, ob die Zugrundelegung einer linearen Modellfunktion überhaupt angebracht ist. Auch wenn sich dies in vielen Fällen praktikabel und vollkommen ausreichend erweist (vgl. S. 29), kann bereits der erste Blick auf das Streudiagramm den Schluss nahelegen, dass eine nicht-lineare Funktion als Modellfunktion eindeutig sinnvoller erscheint (vgl. Abb. 2-19). Aber selbst bei einer linearen Anpassung kann es vorkommen, dass eine zunehmende Streuung der Datenpunkte übersehen wird (vgl. VOGEL & EICHLER 2010a, S. 10).

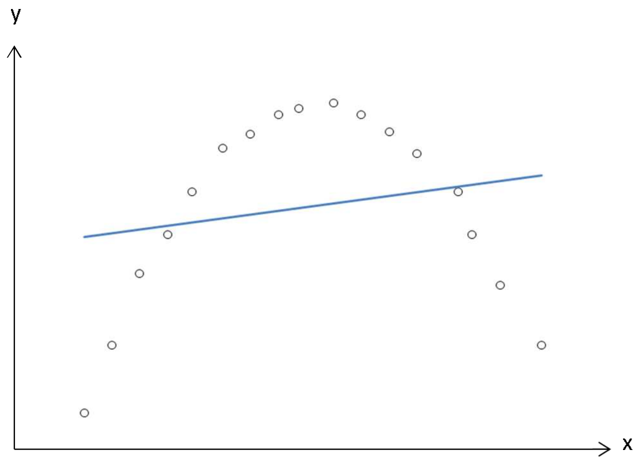


Abb. 2-19: Lineare Ausgleichsfunktion bei offensichtlich quadratischem Zusammenhang?

Eine einfache, manchmal unterschätzte Methode (vgl. VOGEL & EICHLER 2010a, S. 10) zur Untersuchung der Anpassungsgüte einer Regressionsfunkton ist die so genannte **Residuenanalyse**: Dabei werden die Residuen, also die vertikal gemessene Entfernung zwischen dem Ausgleichsgraphen und dem jeweiligen Datenpunkt, in einem separaten Streudiagramm dargestellt. Die jeweiligen Regressionswerte $\hat{y}_i$ (i = 1, ..., n) werden auf der Abszisse, die Residuen auf der Ordinate abgetragen. Ist die Regressionsfunktion passend gewählt, so streuen die Punkte in dem so entstandenen Residuendiagramm relativ gleichmäßig und ohne erkennbare Struktur um die Nulllinie (Abb. 2-20 a, S. 35). Lässt sich im Residuenplot ein (neues) Muster erkennen (Abb. 2-20 b, S. 35), gilt es, „das lineare Modell zu verwerfen oder [...] Ursachen für systematische Abweichungen zu erforschen" (BIEHLER & SCHWEYNOCH 1999, S. 20). Wird die Entscheidung getroffen, dass der

Zusammenhang durch einen anderen Funktionstyp *besser* dargestellt werden könnte, bietet sich dazu nahezu jede Funktionsklasse an (Parabel, Potenz- oder Exponentialfunktion, ...). Das Prinzip der Methode der Kleinsten Quadrate lässt sich dabei auch auf eine nicht-lineare Funktionen übertragen, allerdings gestaltet sich in den allermeisten Fällen die Bestimmung der entsprechenden Parameter der Regressionsfunktion als sehr rechenaufwändig und ist nur noch mithilfe entsprechender Software möglich.

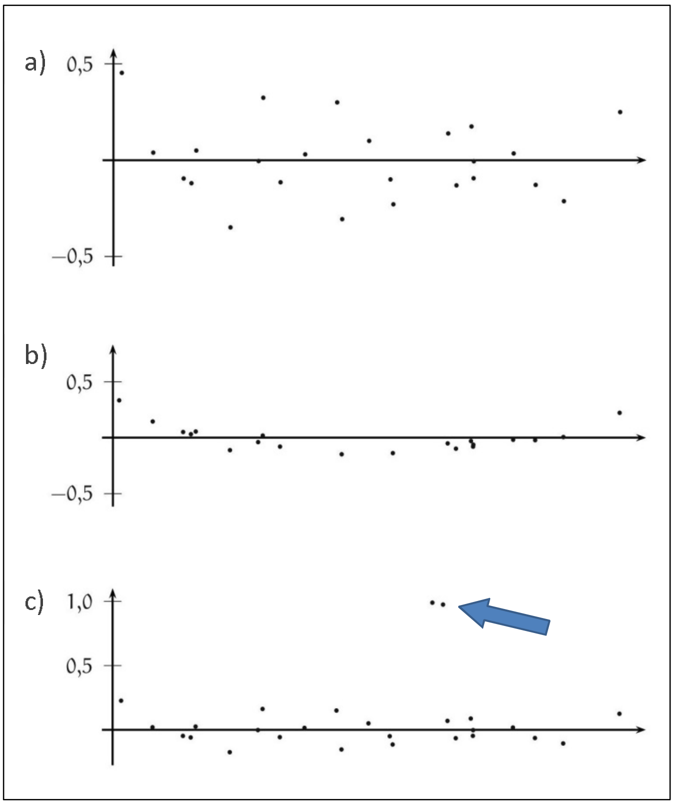


Abb. 2-20: Verschiedene Residuenplots (nach CRAMER & KAMPS 2014, S. 134 f)

Unabhängig vom gewählten Funktionstyp kann ein Residuendiagramm auch bei der Optimierung der gewählten Modellfunktion hilfreich sein (vgl. VOGEL & EICHLER 2010a, S. 11): Ausreißer fallen im Residuenplot für gewöhnlich schneller ins Auge (Abb. 2-20 c) und sollten grundsätzlich genauer betrachtet werden. Unter Umständen können diese aus dem Datensatz entfernt werden (z. B. aufgrund von Messfehlern oder offensichtlichen Falschangaben) und ermöglichen so eine weitere Optimierung der Regressionsfunktion. „Dabei ist zu beachten, dass auch Ausreißer relevante Information enthalten können. Eine entsprechende Bereinigung des Datensatzes ist daher sorgfältig zu rechtfertigen.“ (CRAMER & KAMPS 2014, S. 135). Ohnehin ist es nicht nur aus mathematischer Sicht „eine spannende Frage, ob man sich die Abweichungen vom Trend erklären kann.“ (BIEHLER & SCHWEYNOCH 1999, S. 21).

2.3 Didaktische Reduktion: Bivariate Daten in der Sekundarstufe I

Die obigen Ausführungen zeigen zwar auf, welche Möglichkeiten die *klassische*, deskriptive Statistik für eine Untersuchung bivariater Daten anbietet, allerdings wird schnell klar, dass vor allem in der Sekundarstufe I die entsprechenden mathematischen Grundlagen für eine derartige Analyse nicht zur Verfügung stehen. Die folgenden beiden Kapitel sollen verdeutlichen, wie ein verständiger Umgang mit bivariaten auch in der Sekundarstufe I erfolgen kann. Der Vorstellung der Explorativen Datenanalyse als mögliches didaktisches Konzept zur (generellen) Behandlung (nicht nur) bivariater Datensätze im Unterricht schließt sich eine beispielhafte Analyse zur Illustration an. Dabei wollen die folgenden Ausführungen nicht als didaktischer Ratgeber verstanden werden, sondern als unterstützende bzw. als ergänzende Grundlage für das in Kap. 2.4 abzuleitende Verständnis einer Datenkompetenz im Umgang mit bivariaten Daten.

2.3.1 Explorative Datenanalyse als didaktisches Konzept für einen kompetenzorientierten Zugang im Umgang mit bivariaten Daten

Im Umgang mit bivariaten Daten verlagert sich der Focus von einer tendenziell deskriptiven Auswertung der Daten, bei der in der Regel die Aufbereitung der Daten mittels Kenngrößen und/oder das Überprüfen von Hypothesen im Vordergrund steht, hin zu einer explorativen Gesinnung (vgl. einführende Bemerkungen des Kapitels 2.2): „Explorativ-orientierte Lernende [...] gehen explorativ durch die Daten, bis ihnen etwas auffällt oder sie ein Muster entdecken." (FRISCHEMEIER 2017, S. 49). Diese Analysemethode folgt keinem bestimmten Schema, denn die „Daten, nicht vorweg formulierte Hypothesen bestimmen das Vorgehen." (WINTERMANTEL & VOGEL 2003, S. 10). „Explorative Datenanalyse ist ein dynamischer Prozess, bei dem ausgehend von den bisherigen Ergebnissen immer wieder neue, überschaubare und leicht handhabbare Schritte zur weiteren Erforschung der Situation unternommen werden." (BOROVCNIK 1987, S. 200). Der stark visuelle Charakter ermöglicht zudem auch jüngeren Schüler/inne/n den Zugang zu einer verständigen Analyse, zumal die mathematischen Anforderung überschaubar gehalten werden können: Das Bestimmen der Funktionsgleichung einer einfachen Anpassungsgeraden durch zwei geeignete Punkte im Streudiagramm oder über drei Medianpunkte ist auch mit dem Methodenrepertoire der Sekundarstufe I zu bewerkstelligen (vgl. das folgende Kapitel 2.3.2).

Nicht erst in den letzten Jahren hat sich die Explorative Datenanalyse (EDA) als neue Richtung in der Statistik etabliert. Im Gegensatz zu den beiden *klassischen* Formen, der deskriptiven und inferenziellen Statistik, die sich im Wesentlichen damit befassen, „Wissenschaftler beim Beschreiben des Datenmaterials, beim Testen von Hypothesen und bei der statistischen ‚Absicherung' von Ergebnissen zu unterstützen" (BIELHER 1997, S. 8), nimmt die EDA neue Funktionen wahr: Sie wird besonders zur Hypothesen*generierung* herangezogen, soll Daten (noch) besser und leichter interpretierbar machen und es ermöglichen, „im Interesse des betrachteten Sachproblems zu vertieften Erkenntnissen zu gelangen." (BIEHLER 1997, S. 8). Das erfordert allerdings, dass man sich über die standardisierten Methoden hinaus mit einem Datensatz beschäftigt: Daten müssen „gedreht und gewendet" (BIEHLER 1997, S. 8) werden, um Besonderheiten und Muster sichtbar zu machen. Dazu werden sie für den Analyseprozess bereits relativ früh visualisiert. Hypothesen werden generiert und überprüft, Modelle angepasst und Residuen analysiert. Ausreißer werden besonders beleuchtet und hinterfragt. Entscheidend für den explorativen Ansatz ist zudem, „weitere Hintergrundinformationen zu beschaffen sowie nach Ursachen und Faktoren für die entdeckten Besonderheiten und Zusammenhänge zu forschen." (NOLL & SCHMIDT 1994, S. 9) Dazu müssen „andere Materialien wie Texte, Bilder, Filmausschnitte, Zeitungsartikel, Interviews etc. [...] herangezogen werden, um zu einem Gesamtbild zu kommen, um der ‚Sache' gerecht zu werden." (BIEHLER 1995, S. 6).

Die Vielzahl und Art dieser *detektivischen* Tätigkeiten haben in der Literatur die Metapher des Datendetektivs hervorgebracht; eine Figur, die auf den amerikanischen Forscher JOHN TUKEY zurückgeht und „seitdem in vielen didaktischen Arbeiten zitiert und ausgestaltet wurde" (EICHLER 2006, S. 2). D. VOGEL hat das Ziel einer Explorativen Datenanalyse in diesem Zusammenhang in einem Interview folgendermaßen beschrieben:

„Der Detektiv ist offen für das Unerwartete. Er will Neues entdecken und Verborgenes ans Licht bringen. Seine Fragen sind: Habe ich etwas übersehen? Wie lassen sich die Fakten erklären? Was fällt auf? Sind die Aussagen plausibel? Genau mit dieser Haltung sollte sich auch der Statistiker den Daten nähern. [...] Allerdings kommt es darauf an, dies bewusst und systematisch zu tun [...] und dabei vorhandene explorative Instrumente zu nutzen, sie zu modifizieren und weitere zu ersinnen. Genau dies will die EDA. Die EDA will die herkömmliche Statistik ergänzen, nicht verdrängen. Sie verändert den Blick und ergänzt das Methodenrepertoire.“

(URL: http://www.brd.nrw.de/lerntreffs/mathe/pages/magazin/leute/vogel.html (Stand: 8.6.2017))

Besonders diese vielfältigen, abwechslungsreichen Tätigkeiten scheinen für die Umsetzung und Förderung der in den KMK geforderten prozess- und inhaltsbezogenen mathematischen Kompetenzen (vgl. Kap. 2.5: „Datenkompetenz, bivariate Daten und Bildungspläne ...“) ideal. NOLL und SCHMIDT (1994, S. 10) fassen zusammen:

„Ein solcher Unterricht

- ermöglicht es, den Umgang mit Daten motivierender und spannender zu gestalten,
- erfordert und ermöglicht offenere Arbeitsweisen; Schülerinnen und Schüler können sich selbst stärker als Experten einbringen; [...]
- betont interpretative und begriffliche Aspekte bei der Anwendung von Mathematik; Mathematik als reines Rechnen tritt in den Hintergrund [...].“

Nicht nur aus mathematikdidaktischer Sicht ergeben sich durch den Ansatz explorativer Datenanalyse „neue Perspektiven für selbstständigkeitsförderndes entdeckendes Lernen“ (BIEHLER & HARTUNG 2006, S. 54). Nicht zuletzt weil es bei dieser Form der Aufarbeitung der Daten keine tatsächlich einzig richtige oder falsche Vorgehensweise gibt, ist diese Herangehensweise „ein interaktiver und iterativer Prozess, in dem Umwege und Irrwege als Lernchance begriffen werden“ (BIEHLER 1997, S. 8). Werden Taschenrechner und/oder Statistiksoftware eingesetzt, ergeben sich

dadurch „zusätzliche Chancen für einen forschenden und entdeckenden Unterricht“ (BIEHLER & WEBER 1995, S. 5).
GNANADESIKAN, KETTERING und TUKEY (1982, S. 37) hoffen darüber hinaus, dass „EDA und insbesondere graphische Methoden [...] das ‚Aha‘!“ in die Statistik zurückbringen.

2.3.2 Beispielhafte Analyse eines Datensatzes und Möglichkeiten der Modellierung in der Sekundarstufe I

Die folgende Exploration soll exemplarisch aufzeigen, wie die Analyse eines bivariaten Datensatzes mit den Mitteln der Sekundarstufe I in einem Unterrichtsvorhaben ablaufen kann. Obwohl eine umfassende – und nicht auf die Mittel der Sekundarstufe I beschränkte – explorative Datenanalyse zwar durchaus „recht komplex“ und eigentlich „nicht einfach zu beschreiben“ (NOLL & SCHMIDT 1994, S. 9) ist, kann die folgende Vorgehensweise zumindest als kleines und im Umfang beschränktes Repertoire an Methoden angesehen werden, das sich für eine einführende Auseinandersetzung mit bivariaten Daten anbietet, jedoch nicht zwingend in der dargestellten Reihenfolge erfolgen und natürlich auch nicht komplett abgearbeitet werden muss bzw. soll. Ein „richtiges“ Behandeln gibt es laut BOROVCNIK ohnehin nicht, vielmehr komme es darauf an, „Besonderheiten des Untersuchungsmaterials deutlich hervortreten zu lassen.“ (BOROVCNIK 1987, S. 196).
Die Grundlage der folgenden Überlegungen bildet die Tabelle in Abbildung 2-21 (S. 40), die die durchschnittliche Herzfrequenz sowie die durchschnittliche Lebenserwartung einiger Tiere darstellt.

	Tier	*Herzfrequenz (durchschnittliche Schläge pro Min.)*	*durchschnittliche Lebenserwar-tung*
1	Ente	325	3
2	Fuchs	100	13
3	Goldhamster	425	2
4	Hauskatze	120	13
5	Huhn	353	6
6	Igel, wach	300	7
7	Kaninchen	215	9
8	Karpfen	60	20
9	Krähe	380	3
10	Löwe	40	23
11	Meerschweinchen	256	7
12	Ringelnatter	32	18
13	Schaf	70	18
14	See-Elefant (a. d. Wasseroberfläche)	60	18
15	Storch	85	16
16	Taube	200	15
17	Wanderratte	355	5
18	Ziege	75	20

Abb. 2-21: Tabelle verschiedener Tiere mit der jeweiligen Herzfrequenz und Lebenserwartung

Jeder *klassische* Statistiker würde an diesem Punkt eine Hypothese erwarten, die es zu überprüfen gilt, oder eine konkrete Fragestellung, der er nachgehen kann. In der EDA hingegen ist dies vorrangig gar nicht erwünscht, denn vielmehr soll ja der Datendetektiv selbst „Fragestellungen entwickeln und Hypothesen aufstellen" (NOLL & SCHMIDT 1994, S. 50; vgl. auch Kap. 2.3.1). Auch Personen mit wenig statistischer Vorerfahrung wird sich bei Anblick dieser Tabelle jedoch sicher die Frage aufdrängen, ob zwischen der Herzfrequenz der Tiere und deren Lebenserwartung ein gewisser Zusammenhang besteht: Sobald „im Datenmaterial mehr als zwei Merkmale enthalten sind, besteht ein unwillkürliches Interesse an Zusammenhängen." (EICHLER & VOGEL 2013, S. 75).

In der abgebildeten Form lässt sich allerdings mit bloßem Auge kaum ein Zusammenhang zwischen den Größen feststellen, deshalb besteht der erste Schritt in der **Umordnung** bzw. (aufsteigenden) **Sortierung der Tabelle** (hier nach der durchschnittlichen Herzfrequenz). Die Abbildung 2-22 (S. 41) zeigt das Ergebnis dieser Neusortierung.

	Tier	*Herzfrequenz (durchschnittliche Schläge pro Min.)*	*durchschnittliche Lebenserwartung*
*1**	Ringelnatter	32	18
*2**	Löwe	40	23
*3**	Karpfen	60	20
*4**	See-Elefant (a. d. Wasseroberfläche)	60	18
*5**	Schaf	70	18
*6**	Ziege	75	20
*7**	Storch	85	16
*8**	Fuchs	100	13
*9**	Hauskatze	120	13
*10**	Taube	200	15
*11**	Kaninchen	215	9
*12**	Meerschweinchen	256	7
*13**	Igel, wach	300	7
*14**	Ente	325	3
*15**	Huhn	353	6
*16**	Wanderratte	355	5
*17**	Krähe	380	3
*18**	Goldhamster	425	2

Abb. 2-22: Nach Herzfrequenz sortierte Tabelle

Mit etwas Übung im Tabellenlesen wird man bereits auf die gegenläufige Tendenz der Werte aufmerksam werden und geneigt sein, sich (vorerst) zu einer Aussage wie „Je höher die Herzfrequenz, desto niedriger die Lebenserwartung des Tieres“ verleiten lassen.

Im nächsten Schritt der Analyse offenbart sich der stark visuelle Charakter der EDA: Um eventuell auftretende Muster und Besonderheiten des Datensatzes erkennen zu können, werden die Daten „graphisch so aufbereitet, dass Besonderheiten durch bloßes Hinsehen erkannt werden können.“ (Borovcnik 1987, S. 197). Generell empfiehlt sich vor allem bei größeren Datensätzen, rasch zu einer graphischen Darstellung überzugehen, denn „sie verspricht mehr Übersicht und damit die größere Chance, Trends oder andere Besonderheiten zu entdecken.“ (Noll & Schmidt 1994, S. 52). Im vorliegenden Beispiel visualisieren wir die Daten mit Hilfe eines **Streudiagramms** (vgl. S. 21).

In der Abbildung 2-23 ist die durchschnittliche Herzfrequenz der Tiere auf der Rechtswertachse (x-Achse) abgetragen, die jeweils dazugehörige durchschnittliche Lebenserwartung auf der Hochwertachse (y-Achse).

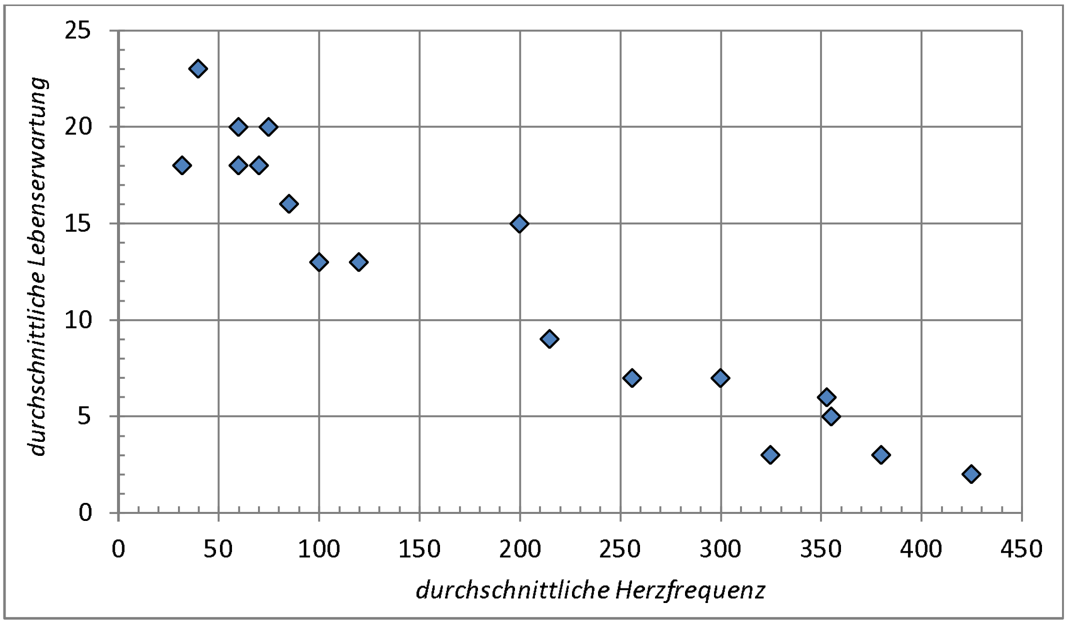


Abb. 2-23: Herzfrequenz (x-Achse) und Lebenserwartung (y-Achse) als Streudiagramm

Die oben geäußerte Zusammenhangshypothese scheint sich in der grafischen Darstellung zu bestätigen. Der Versuch, den Zusammenhang durch eine Gerade darzustellen, scheint nicht abwegig. Eine solche **Anpassungsgerade** oder Ausgleichsgerade muss nun für eine erste Annäherung gerade in der Sekundarstufe I nicht zwingend rechnerisch ermittelt werden; für eine erste Einpassung genügt es, „sich an der Steigung der Punktbahn zu orientieren und dann eine mittlere Position in dem Band zu wählen" (BIEHLER & SCHWEYNOCH 1999, S. 19). Für eine anschließende Bestimmung der Funktionsgleichung jedoch ist es hilfreich, wenn die per Augenmaß eingepasste Gerade zumindest durch zwei eingezeichnete Punkte verläuft (siehe Abb. 2-24, S. 43).

Vor allem in den höheren Jahrgangsstufen der Sekundarstufe I ist die Ermittlung der Funktionsgleichung einer solchen Anpassungsgeraden ohne Probleme möglich.

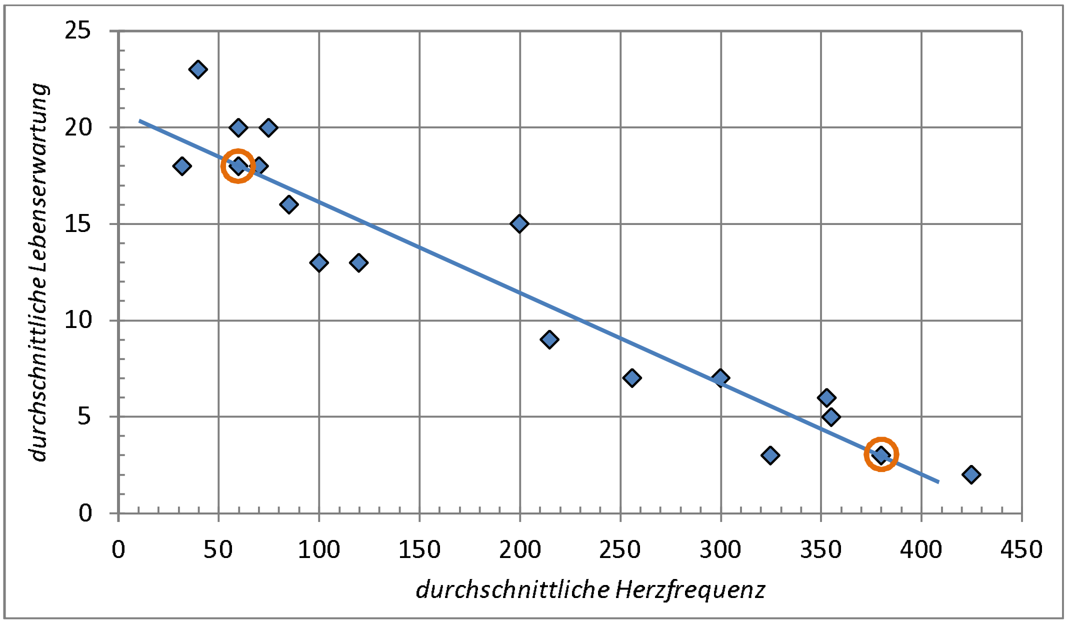


Abb. 2-24: Das Streudiagramm mit einer ersten Ausgleichsgeraden („Trendgerade“)

Die in Abbildung 2-24 durch die Punkte (60 | 18) und (380 | 3) eingezeichnete Ausgleichsgerade ist rechnerisch leicht zu ermitteln: $\Delta x = 380 - 60 = 320$

$\Delta y = 3 - 18 = -15$

$m = \frac{\Delta y}{\Delta x} = \frac{-15}{320} = -0{,}046875$

Bestimmen des y-Achsenabschnitts t: $18 = -0{,}046875 \cdot 60 + t$

$\Leftrightarrow$ $t = 20{,}8125$

$\rightarrow$ <u>$y = -0{,}046875x + 20{,}8125$</u>

Diese Ausgleichsgerade mittelt nun alle Wertepaare (Herzfrequenz | Lebenserwartung), so dass für jede Merkmalsausprägung des einen Merkmals die modellhaft angenommene Merkmalsausprägung des zweiten Merkmals angegeben werden kann. Der dadurch erkennbare Trend ermutigt uns einerseits, die anfangs ausgesprochene Hypothese beizubehalten (je höher die Herzfrequenz, desto niedriger die Lebenserwartung des Tieres), andererseits kann er „als Hilfe für Prognosen verwendet werden“ (NOLL & SCHMIDT 1994, S. 31): Welche (durchschnittliche) Lebenserwartung würde man bei einem Tier erwarten, dessen Herz durchschnittlich 200 Mal pro Minute schlägt?

Auch in der Sekundarstufe I und selbst bei der bloßen Einpassung per Augenmaß ist die Betrachtung der Residuen unerlässlich (vgl. S. 34), denn sie gibt nun im Folgenden Aufschluss darüber, wie *gut* die Gerade eingepasst wurde und inwieweit nachgebessert werden sollte.
Obwohl die per Augenmaß eingepassten Geraden in der Praxis häufig nur gering von rechnerisch ermittelten Ausgleichsgeraden abweichen (vgl. Biehler & Schweynoch 1999, S. 21 und Stepancik 2012, S. 21), gibt es weitere, auch für den Einsatz in der Sekundarstufe I geeignete Verfahren, um die vermeintliche ideale Anpassungsgerade zu bestimmen. Auf die zwei bekanntesten und am häufigsten verwendeten wird im Folgenden kurz eingegangen.

Die **Median-Median-Gerade** (auch 3-Gruppen-Gerade bzw. 3-Schnitt-Median-Gerade genannt) lässt sich noch mit dem Methodenrepertoire der Sekundarstufe I bewerkstelligen: Dazu werden die ursprünglichen Daten in drei Gruppen (Cluster) unterteilt und die jeweiligen Mediane bestimmt (siehe Abb. 2-25).

	Tier	***Herzfrequenz (durchschnittliche Schläge pro Min.)***	***durchschnittliche Lebenserwartung***	
*1**	Ringelnatter	32	18	
*2**	Löwe	40	23	
*3**	Karpfen	60	20	$M_1(60\|19)$
*4**	See-Elefant	60	18	
*5**	Schaf	70	18	
*6**	Ziege	75	20	
*7**	Storch	85	16	
*8**	Fuchs	100	13	
*9**	Hauskatze	120	13	$M_2(160\|13)$
*10**	Taube	200	15	
*11**	Kaninchen	215	9	
*12**	Meerschweinchen	256	7	
*13**	Igel, wach	300	7	
*14**	Ente	325	3	
*15**	Huhn	353	6	$M_3(354\|4)$
*16**	Wanderratte	355	5	
*17**	Krähe	380	3	
*18**	Goldhamster	425	2	

Abb. 2-25: Die bereits sortierte Tabelle mit den drei Medianen M_1, M_2 und M_3.

Durch diese drei Medianpunkte werden zueinander parallele Geraden gelegt, wobei sich die Steigung m dieser drei Geraden anschließend aus den Koordinaten des ersten und dritten Medians berechnet:

$M_1(60 \mid 19)$, $M_3(354 \mid 4)$ $\quad\rightarrow\quad$ $\Delta x = x_{M3} - x_{M1} = 354 - 60 = 294$

$\Delta y = y_{M3} - y_{M1} = 4 - 19 = -15$

$m = \frac{-15}{294} \approx -0{,}05102$

Der y-Achsenabschnitt t der gesuchten Ausgleichsgeraden ist das arithmetische Mittel aus t_1, t_2 und t_3, wobei natürlich t_1 und t_3 identisch sind.

$$\bar{t} = \frac{t_1 + t_2 + t_3}{3} = \frac{22{,}06 + 21{,}16 + 22{,}06}{3} = \frac{65{,}28}{3} = 21{,}76$$

Berechnung von t_1, t_2 und t_3: $\quad (y = m \cdot x + t)$

$M_3(354 \mid 4)$: $\quad 4 = -0{,}05102 \cdot 354 + t_3 \quad \Leftrightarrow \quad t_3 \approx 22{,}06 \quad (= t_1)$

$M_2(160 \mid 13)$: $\quad 13 = -0{,}05102 \cdot 160 + t_2 \quad \Leftrightarrow \quad t_2 \approx 21{,}16$

Somit ergibt sich als Funktionsgleichung für diese Median-Median-Gerade
$y = -0{,}05102 + 21{,}76$.

Die Abbildung 2-26 illustriert die errechneten Median-Punkte inklusive der Geraden durch diese Punkte sowie die eben errechnete Median-Median-Gerade.

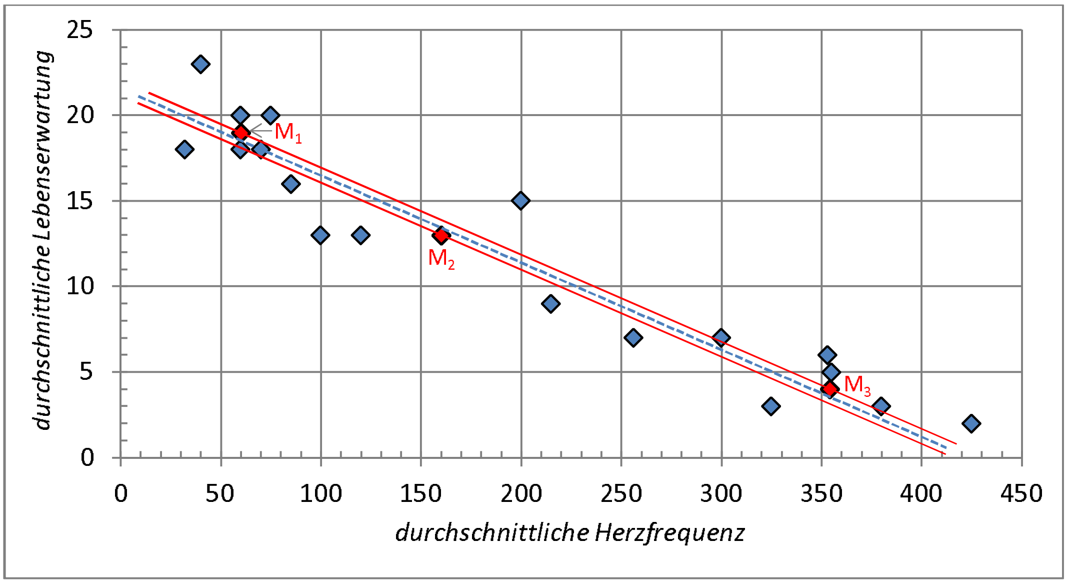


Abb. 2-26: Das Streudiagramm inkl. der drei Medianpunkte, der vorläufigen parallelen Geraden durch M_1, M_2 und M_3 (rot) sowie der errechneten Ausgleichsgeraden mit $\bar{t}$ (gestrichelte Linie, blau)

Der y-Achsenabschnitt $\bar{t}$ hätte nicht unbedingt über das arithmetische Mittel errechnet werden müssen, Noll und Schmidt weisen auf die Möglichkeit hin, dass eine reine Parallelverschiebung der Geraden, so dass sie „möglichst genau zwischen den Medianpunkten liegt“ (Noll & Schmidt 1994, S. 40), auch genügen kann. Die bereits in Kapitel 2.2.7 dargestellte Gaußsche Fehlerquadratmethode (siehe S. 28 ff) gilt als ein Standardverfahren der Statistik, bei der versucht wird, eine Funktion zu finden, bei der die Quadrate der Residuen in der Summe minimal sind. Die Abbildung 2-27 veranschaulicht die Residuenquadrate einiger (ausgewählter) Punkte.

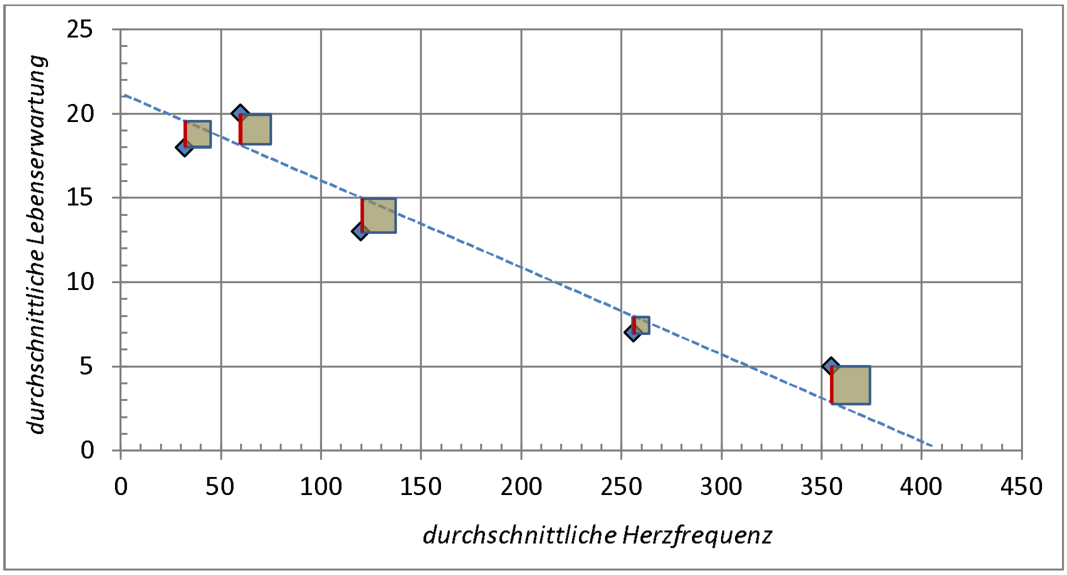


Abb. 2-27: Anschauliche Darstellung der Residuenquadrate an ausgewählten Punkten

Die Anpassungsfunktion muss dabei nicht wie in diesem Beispiel eine Gerade sein; fast jede Statistiksoftware bietet als Regressionsmodell eine lineare, logistische, exponentielle oder eine Polynom- oder Potenzfunktion an. Die (Rechen-) Verfahren dazu sind aufwändig und dürfen aus didaktischen Gründen auch in der Sekundarstufe II höherwertigen Taschenrechnern oder einer Statistiksoftware überlassen werden (eine Formel zur Berechnung findet sich z. B. bei Eichler & Vogel 2013, S. 127). Die dahinter steckende Idee sollte und kann jedoch durchaus im Unterricht thematisiert werden, denn gerade die automatisierte Berechnung auf Knopfdruck kann eine „unkritische Anwendung dieser Methode“ (Noll & Schmidt 1994, S. 43) begünstigen, was einem entscheidenden Grundanliegen der EDA zuwiderlaufen würde. NOLL und SCHMIDT geben hier den Rat: „Berechne die Ausgleichsgerade nie ohne vorherige kritische Überprüfung der graphischen Darstellung [...]!“ (NOLL & SCHMIDT 1994, S. 43). Im vorliegenden Beispiel ermittelte die Software GeoGebra (ebenso wie beispielsweise Excel 2016) die Geradengleichung

$$y = 0{,}0478x + 21{,}1589$$

als Ausgleichsgerade über die Gaußsche Fehlerquadratmethode. Sofern sich im Unterricht überhaupt die Frage stellt, welche der ermittelten Geraden nun die *bessere* sei, ist zu überlegen, ob es grundätzlich ratsam ist, in der Sekundarstufe I verschiedene Verfahren zur Ermittlung der Ausgleichsgeraden aufzuzeigen. EICHLER und VOGEL raten, darüber nachzudenken, ob „die Frage nach der *besten* Geraden überhaupt thematisiert werden soll“ (EICHLER & VOGEL 2013, S. 96). Auch sie halten das Prinzip der Methode der kleinsten Quadrate vor allem aus didaktischer Sicht „nicht für die Sekundarstufe I geeignet“ (EICHLER & VOGEL 2013, S. 96). Das Einpassen nach Augenmaß hingegen überrascht (vor allem nach etwas Übung) mit erstaunlicher Treffsicherheit in Bezug auf rechnerische Anpassungen und kann bereits ab der 5. Jahrgangsstufe angewendet werden.

Ein wichtiges Hilfsmittel, das neben den oben angeführten Überlegungen Aufschluss über die Güte der Ausgleichsgeraden geben kann, ist das **Residuendiagramm** (vgl. Kap. 2.2.8): Dabei werden die Residuen in einem separaten Diagramm in Bezug zur ermittelten Modellfunktion (Anpassungsgerade) dargestellt.

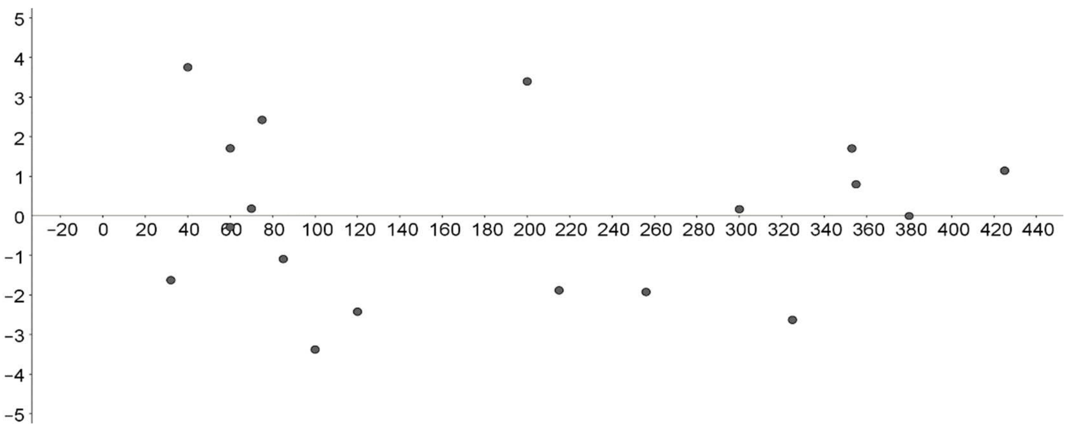


Abb. 2-28: Residuenplot zur kQ-Geraden

Die in Abbildung 2-28 (S. 47) dargestellten Residuen streuen relativ zufällig und ohne erkennbares Muster um die Nulllinie, zudem verteilen sie sich beinahe gleichmäßig in die obere und untere Hälfte; es kann also die Annahme eines linearen Zusammenhangs beibehalten werden. Ansonsten müsste man „das lineare Modell [...] verwerfen oder [...] Ursachen für systematische Abweichungen [...] erforschen“ (BIEHLER & SCHWEYNOCH 1999, S. 20). Generell lässt eine systematische Residuenabweichung darauf schließen, dass die gewählte Modellfunktion ihrer Erklärkraft nicht standhält (vgl. VOGEL & EICHLER 2010a, S. 10).

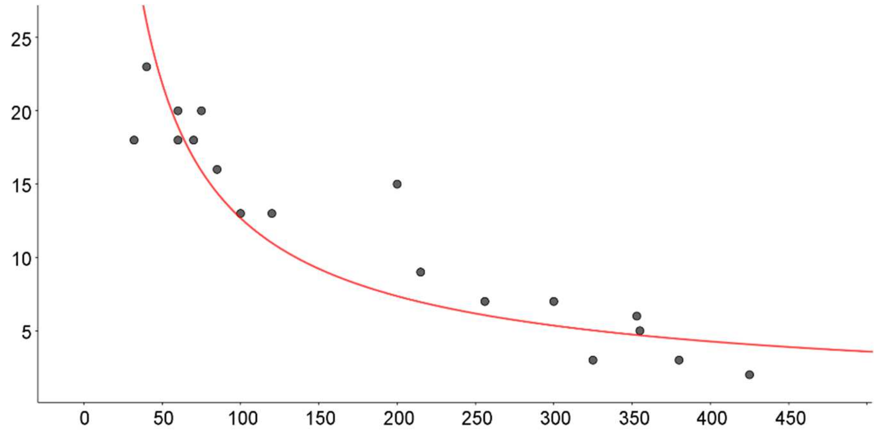
Trotz des vermeintlich bzw. vorerst zufriedenstellenden Ergebnisses bleibt der Datendetektiv auch weiterhin seiner explorativen Gesinnung treu: Fragen, die sich auch bereits vor der Analyse hätten stellen lassen, verlieren auch nach einer Visualisierung und Interpretation nicht an Bedeutung (vgl. Kap. 2.3.1): „Das Gewinnen von Erkenntnissen über außermathematische Zusammenhänge [...] ist wesentlicher Bestandteil der Explorativen Datenanalyse.“ (NOLL & SCHMIDT 1995, S. 50). Überlegungen zur **Plausibilität** des entdeckten Zusammenhangs dürfen somit auch hier keinesfalls vernachlässigt werden.

Auch wenn der im aufgezeigten Beispiel der entdeckte Zusammenhang plausibel erscheint, bleiben doch einige, nicht unwichtige Fragen offen: Woher kommen die Daten, wer hat sie erhoben bzw. recherchiert? Wie *zuverlässig* sind sie?

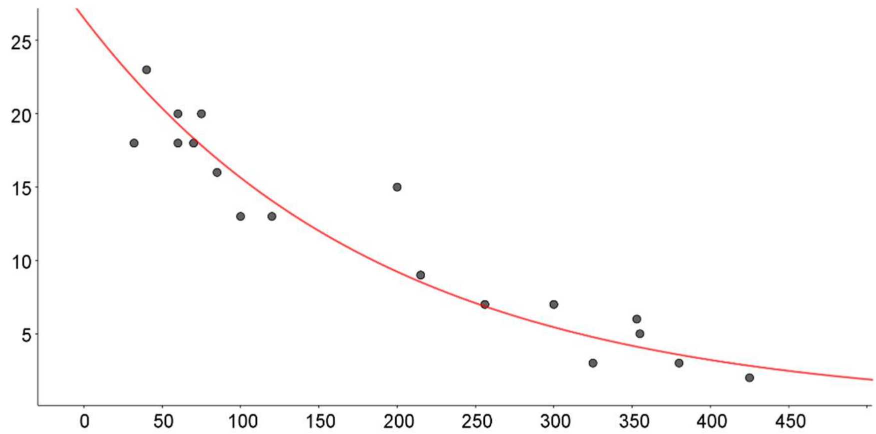
Wer selbst (nach-) recherchiert und versucht, in Lexika oder im WWW das Durchschnittsalter und den Ruhepuls gewisser Tiere zu dokumentieren, stellt rasch fest, dass in der Regel ohnehin bereits nur Durchschnittswerte angegeben sind. Offenbar gibt es zudem große Unterschiede, ob es sich um Haustiere oder Tiere in freier Wildbahn handelt, vor allem was die Angaben einer möglichen Lebenserwartung betrifft. Weitere Fragen drängen sich auf: Müsste man nicht nach Geschlecht differenzieren? Hat das Alter der Tiere einen Einfluss auf den (Ruhe-) Puls? Wäre es nicht sinnvoll, nach Tierarten zu clustern, zum Beispiel in Säugetiere oder wirbellose Tiere?

Ist der Datensatz groß genug? Ergibt sich ein anderes Bild bzw. eine neue Erkenntnis, wenn uns weitere Daten zur Verfügung stünden? Selbstverständlich unterliegen Puls und Lebenserwartung natürlichen Grenzen – kann dann der entdeckte Zusammenhang wirklich linear sein? Fragen dieser Art sollten auch in der

Sekundarstufe I thematisiert werden; zumindest die Demonstration möglicher nicht-linearer Anpassungsfunktionen kann den Modellierungsgedanken weiter unterstützen. Alleine durch das Aufzeigen der grundsätzlichen Möglichkeiten wird der Horizont der Schüler/innen erweitert und damit die kritische Grundeinstellung gegenüber statistischen Aussagen gefördert.
Entsprechende Programme beinhalten als Tool weitere Regressionsfunktionen (vgl. S. 47), die über wenige Mausklicks auch Anpassungskurven liefern, die vor allem dem Aspekt der Begrenztheit genügen. Die beiden Streudiagramme (A) und (B) in Abbildung 2-29 demonstrieren zwei nicht-lineare Regressionsfunktionen und bestätigen den nicht explizit ausgesprochenen Anfangsverdacht, dass der Zusammenhang zwischen durchschnittlicher Herzfrequenz und Lebenserwartung durch ein Kurve offenbar besser dargestellt werden als durch eine Gerade.



(A) Anpassung über eine Potenzfunktion: $y = 471{,}5663 \cdot x^{-0{,}7852}$



(B) Anpassung über eine Exponentialfunktion: $y = 26{,}5094 \cdot 0{,}9947^x$

Abb. 2-29: Zwei Beispiele für nicht-lineare Modelle der Zusammenhangshypothese, ermittelt mit Hilfe der Software GeoGebra

Darüber hinaus kann gegebenenfalls überprüft werden, wie sich eine gewählte Modellfunktion verhält, wenn man den Datensatz vergrößert (vgl. Abb. 2-30). Die Abbildung 2-31 (S. 51) zeigt im Anschluss eine Regressionsfunktion zu diesem erweiterten Datensatz (Potenzfunktion).

	Tier	*Herzfrequenz (durchschnittliche Schläge pro Min.)*	*durchschnittliche Lebenserwartung*
1	Delphin	99	28
2	Ente	325	3
3	Erdkröte	45	11
4	Fuchs	100	13
5	Gans	80	5
6	Giraffe	66	25
7	Goldfisch	38	30
8	Goldhamster	425	2
9	Hauskatze	120	13
10	Huhn	353	6
11	Hund	120	23
12	Igel, wach	300	7
13	Kaninchen	215	9
14	Karpfen	60	20
15	Krähe	380	3
16	Kreuzotter	40	15
17	Löwe	40	23
18	Maus	500	3
19	Meerschweinchen	256	7
20	Panther	60	13
21	Pferd	38	35
22	Rind	48	26
23	Ringelnatter	32	18
24	Schaf	70	18
25	Schwein	70	11
26	See-Elefant (an der Wasseroberflä-	60	18
27	Storch	85	16
28	Strauß	65	35
29	Taube	200	15
30	Wal	16	39
31	Wanderratte	355	5
32	Ziege	75	20

Abb. 2-30: Erweiterter (unsortierter) Datensatz für eine mögliche Verifizierung der gewählten Funktionsklasse der Modellfunktion

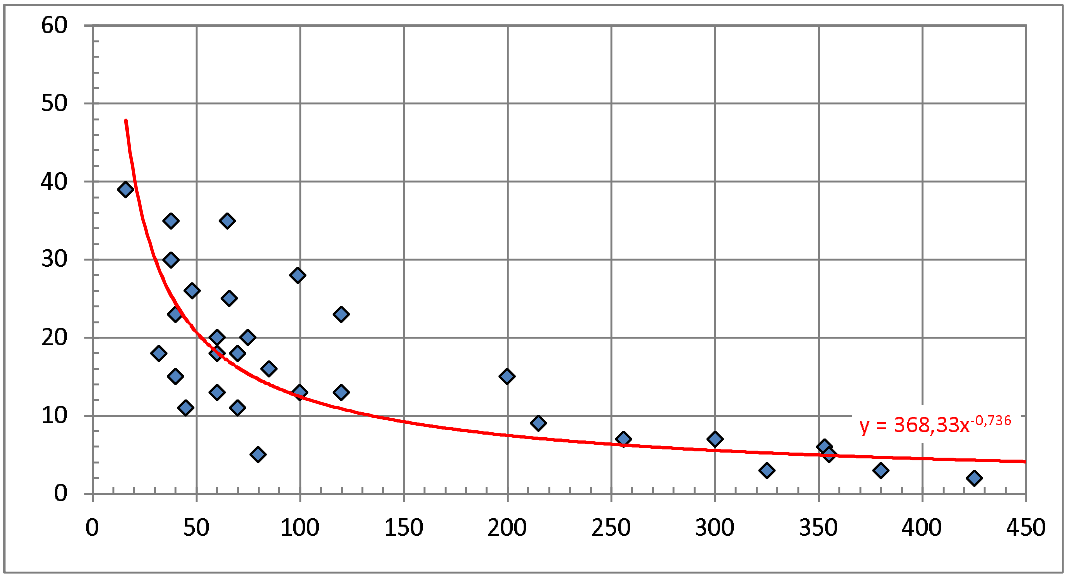


Abb. 2-31: Modellfunktion aus der Klasse der Potenzfunktionen zum erweiterter Datensatz, ermittelt mit Excel 2016

Die Erweiterung des Datensatzes muss indes keineswegs das Anpassungsmodell so wohlwollend stützen wie es in diesem Beispiel der Fall ist. Zusätzliche neue Daten können ein Modell unter Umständen auch ganz zu Fall bringen, was dazu führt, dass der Modellierungskreislauf erneut begonnen werden muss – was den forschenden und suchenden Charakter einer explorativen Datenanalyse erneut unterstreicht.

Zudem wird an dieser Stelle deutlich, wie wichtig der Aspekt des Erkennens der Notwendigkeit von (ausreichend vielen) Daten („*recognition of the need for data*", einer der fünf Säulen statistischen Denkens nach WILD & PFANNKUCH, vgl. S. 10) für die Datenanalyse ist: Die Einsicht, dass Daten notwendig sind, um Überzeugungen oder Vermutungen statistisch zu untermauern, ist ein Prozess, der unabdingbar mit Datenkompetenz verbunden ist und sich im Laufe der Zeit erst entwickeln muss (vgl. EICHLER & VOGEL 2013, S. XII). Auch hier kann die EDA einen wertvollen Beitrag leisten.

Auch wenn nicht-lineare Anpassungsmodelle wie in den Abbildungen 2-29 und 2-31 in der Didaktik nicht unkritisch betrachtet werden, da sie ohne Softwareunterstützung kaum zu bewerkstelligen sind (vgl. Bemerkungen auf S. 46 f), plädieren manche Didaktiker genau aus diesem Grund für den Einsatz solcher Software: Die Analyse und Optimierung der erzeugten Modelle stehe im Vordergrund, nicht das Errechnen der notwendigen Variablen (vgl. STEPANCIK 2012, S. 21). In diesem Sinne kann ein Blick über den Tellerrand hinaus daher durchaus auch in der Sekundarstufe I legitimiert werden.

Das Erwähnen der Möglichkeit einer nicht-linearen Anpassung auch in der Sekundarstufe I sollte im obigen Beispiel alleine schon deswegen thematisiert werden, weil sie die natürlichen Grenzen von Alter und Herzfrequenz berücksichtigt. Aufgeweckte Schüler/innen werden ohnehin bei einer rein linearen Anpassung sicherlich von sich aus auf die Nullstellenproblematik zu sprechen kommen.

2.4 Die Sicht dieser Studie auf Datenkompetenz im Umgang mit bivariaten Daten: Ein hierarchisches Modell als Arbeitsdefinition

Die Sachanalyse, die beispielhafte Exploration eines bivariaten Datensatzes und die grundsätzlichen Ausführungen zur EDA haben verdeutlicht, welche hohe Bedeutung besonders der Aspekt des Lesens und Interpretierens von Daten – gleichgültig ob uni- oder bivariat – im Zusammenhang mit Datenkompetenz hat. Speziell bei einer bivariaten Datenanalyse kommt der stark visuelle Charakter (vgl. BOROVCNIK 1987, S. 197) noch deutlicher zum Ausdruck. Das Lesen und Interpretieren statistischer Daten kann damit als eine Schlüsselqualifikation bzw. Kernkompetenz im Umgang mit bivariaten Daten angesehen werden. Das dieser Arbeit zugrunde liegende Verständnis von „Datenkompetenz im Umgang mit bivariaten Daten" wird im weiteren Verlauf somit als das verständige Lesen und Interpretieren von Daten in grafischen oder tabellarischen Darstellungen aufgefasst, durch welches letztendlich das Entdecken eventuell vorhandener Muster oder eines Trends ermöglicht wird.

Dazu wird hauptsächlich auf das ursprünglich von CURCIO (bereits 1981) entwickelte dreistufige Modell der Kompetenz im Umgang mit grafischen Darstellungen zurückgegriffen, das sich im Laufe der Jahre zu einem der bekanntesten und am weitesten verbreitetsten Modelle entwickelte (vgl. GONZALEZ, ESPINEL & AINLEY 2011, S. 189). Diese drei Kompetenzstufen

① *reading the data*,

② *reading between the data* und

③ *reading beyond the data*

folgen einer empirisch überprüften Hierarchie (vgl. Shaugnessy 2007, p. 989 f), bei der die höchste Stufe (*reading beyond the data*) genau jene Fähigkeiten beschreibt, die in dieser Studie mit untersucht werden soll: Das Erfassen eines Trends bzw. eines Musters in den Daten, um eine Vorhersage für weitere Merkmalspaare treffen zu können; genauer: um für jede Merkmalsausprägung des einen Merkmals die modellhaft angenommene Merkmalsausprägung des zweiten Merkmals angeben zu können.
„Die [drei] Abstufungen können zur Erfassung der Verstehensvoraussetzungen des Betrachters genutzt werden. [...] Denn die Fähigkeit eine [...] Darstellung zu verstehen, hängt davon ab, welche Bedeutung der Betrachter auf der jeweiligen Ebene [...] ableiten kann." (SCHERRMANN 2013, S. 166 f). Im Folgenden werden deshalb diese drei Abstufungen konkreter beschrieben.

① *Reading the data:* Das reine *Lesen* der Daten beschreibt auf der untersten Stufe das tatsächliche *Herauslesen* der Informationen, die beispielsweise in einem Diagramm oder einer Tabelle enthalten sind. Fragen können direkt durch das Entnehmen bzw. Lesen des entsprechenden (zugeordneten) Wertes beantwortet werden. (z. B. „Wie viel Taschengeld erhält ein zwölfjähriger Schüler pro Woche?"). Dazu gehört nicht nur das Ablesen einzelner Werte, sondern auch das Erfassen des Titels bzw. des Themas und der Rubrikenachsen bzw. Spaltenbezeichnungen bei Tabellen (z. B. „Was thematisiert diese Grafik?", „Welche Größen werden hier in einen Zusammenhang gebracht?").

② *Reading between the data*: Auf der zweiten Stufe ist es dem Betrachter des Diagramms oder der Tabelle möglich, „Vergleiche *zwischen* den Teilelementen der [...] Darstellung anzustellen" (SCHERRMANN 2013, S. 166). Darüber hinaus sind ggf. einfache Rechenoperation nötig, um entsprechende Fragestellungen auf diesem Level sachgerecht beantworten zu können: *„This includes making comparisons (e. g. greater than, greatest, tallest, smallest, etc.) as well as applying operations."*

(FRIEL & BRIGHT 1995, p. 4). Ausgehend von einer grafischen Darstellung wie in Abbildung 2-32 dargestellt, können auf dieser Ebene Fragen wie „Wie alt ist der Schüler, der am meisten Taschengeld erhält?“ oder aber auch „Bekommt der Zwölfjährige tatsächlich doppelt so viel Taschengeld wie der Sechsjährige?“ korrekt beantwortet werden.

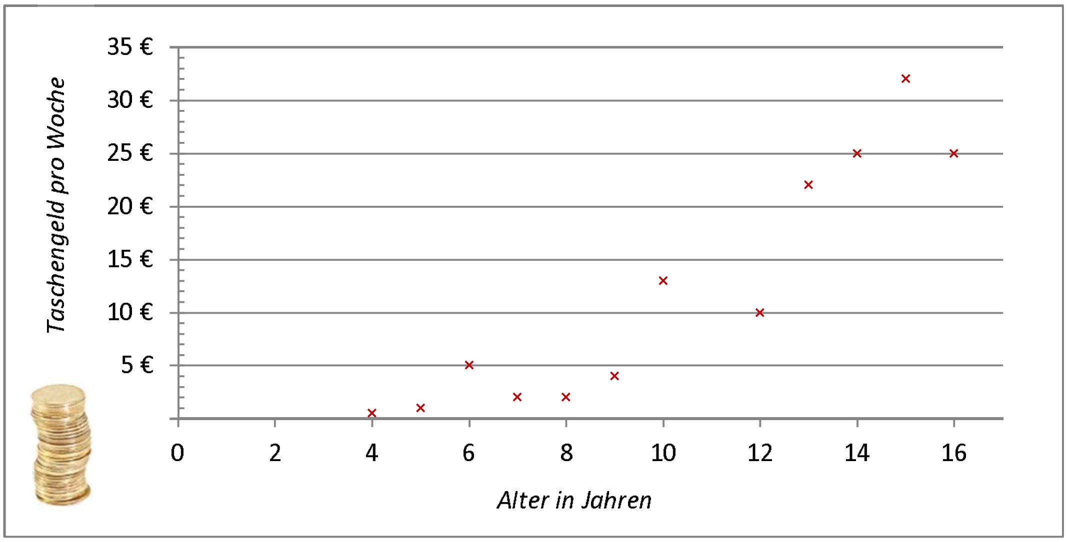


Abb. 2-32: Grafische Darstellung der Taschengeldhöhe in Abhängigkeit vom Alter

③ *Reading beyond the data:* „Auf der elaboriertesten Stufe schließlich kann der Betrachter über die jeweilige grafische Darstellung *hinausreichende* Vorhersagen treffen oder Schlussfolgerungen ziehen“ (SCHERRMANN 2013, S. 166). Damit ist aber nicht sogleich und ausschließlich das Erfassen eines Trends gemeint, sondern auch das korrekte Schlussfolgern und Beantworten von Fragestellungen, die sich so nicht direkt aus der Darstellung entnehmen lassen (z. B. „Wenn der Zehnjährige künftig doppelt so viel Taschengeld pro Woche bekäme und der Vierzehnjährige fünf Euro pro Woche mehr, wer würde dann mehr Taschengeld bekommen?“). Dennoch sind die Einordnung hypothetischer Werte in ein bestehendes Datengefüge („Was vermutest du, wie viel Taschengeld ein elfjähriger Schüler pro Woche erhält?“) und das Erkennen eines Trends meist typische Fragstellungen, um die Zugehörigkeit zu dieser dritten und höchsten Kompetenzstufe zu validieren: *„Students must make inferences from the representation in order to [...] make a prediction about an unknown case [...] or to identify a trend.“* (FRIEL, CURCIO & BRIGHT 2001, p. 132).

Auf dieser dritten Ebene lässt sich auch eine für diese Arbeit entscheidende Auffassung der fünf zentralen Aspekte statistischen Denkens nach WILD und PFANNKUCH einordnen, die zu Beginn der vorliegenden Arbeit im Kapitel 2.1 bei der Begriffseingrenzung einer allgemeinen Datenkompetenz vorgestellt wurden (vgl. S. 10). Einer der Aspekte umschreibt mit *A distinctive set of models*, dass Datenkompetenz das „Suchen, Identifizieren und Beschreiben von Mustern in den Daten“ (VOGEL & EICHLER 2010b, S. 882) umfasst.

Auch wenn dieses dreistufige Modell und die Diskussion darum vorrangig am Verständnis grafischer Darstellungen entfaltet wird, lässt sich diese Einteilung durchaus auch auf das Verständnis tabellarisch dargestellten Daten erweitern. Da in der Literatur explizit kaum darauf eingegangen wird, soll an dieser Stelle das Modell konkret um drei entsprechende Ebenen im Umgang mit Daten in tabellarischen Darstellungen analog zu den von CURCIO gewählten Formulierungen weiterentwickelt werden:

① Das punktuelle Entnehmen und *Ablesen* einzelner Werte oder Wertepaaren aus einer Tabelle ⟶ *reading the data*

(Die Antwort auf eine Frage befindet sich in aller Regel in der direkt benachbarten Zelle der Tabelle.)

---

② Das Vergleichen und Bewerten von Werten und Wertepaaren *innerhalb* einer Tabelle ⟶ *reading between the data*

(Das erfordert ein Springen *zwischen den Zeilen.)*

---

③ Das Schlussfolgern und Vorhersagen gewisser Wertepaare *über* den bestehenden Datensatz *hinaus* ⟶ *reading beyond the data*

(*Dies entspräche dem Einfügen einer neuen Tabellenzeile zwischen den bereits vorhandenen Zeilen.)*

Gleichgültig, ob Daten tabellarisch vorliegen oder in einer Grafik veranschaulicht sind: Jede Statistik ist in einen Kontext eingebunden. Gerade bei der Beurteilung von vermeintlichen Zusammenhängen und Trends (*reading beyond the data*) darf der Sachkontext nicht außer Acht gelassen werden. SHAUGHNESSY hat deswegen diesem Modell bereits relativ früh eine vierte Ebene hinzugefügt: *Reading* behind *the data* („*I previously argued [...] for another level [...] beyond Curcio's three levels, that is, reading* behind *the data or graph.*“, SHAUGHNESSY 2007, p. 989). Was *hinter* den dargebotenen Daten steckt, erfordert selbstverständlich eine weitergehende Recherche über den vorliegenden Datensatz hinaus. Historische, wirtschaftliche oder demografische Einflussfaktoren als mögliche Erklärungsansätze für Unregelmäßigkeiten im Datensatz heranzuziehen – oder einfach nur um zu eruieren, dass man nicht dem Irrtum einer Scheinkausalität erliegt – ist definitiv ein entscheidender Bestandteil von Datenkompetenz. Auch WILD und PFANNKUCH postulieren in diesem Zusammenhang die Integration des Sachkontexts („*One cannot indulge in statistical thinking without some context knowledge*“, WILD & PFANNKUCH 1999, p. 228) als eine ihrer fünf Säulen statistischen Denkens („*Context knowledge, statistical knowledge and synthesis*“, vgl. Kap. 2.1, S. 10).

Diese vierte und sicher ganz wesentliche Komponente einer Datenkompetenz ist allerdings – entgegen der Auffassung SCHERRMANNS, die diese vier Stufen „zum fortschreitenden Verstehen grafischer Veranschaulichungen“ (SCHERRMANN 2013, S. 168) mit entsprechenden Vorwissenskomponenten verknüpft – nicht zwangsläufig als höchste Stufe zu verstehen, sondern wohl eher als ergänzende Komponente anzusehen. Sie ist als vierte Stufe auch nicht empirisch belegt, auch wenn sie gemäß den Abläufen einer Datenanalyse in der Regel an letzter Stelle steht: Für die Beurteilung, welches Merkmalspaar als Ausreißer gilt, muss in der Regel erst eine Modellfunktion eingepasst bzw. ein Trend identifiziert worden sein. Allerdings sind in einem (sinnvoll skalierten) Streudiagramm auch ohne eingezeichnete Regressionsfunktion Besonderheiten im Datenmaterial deutlich sichtbar. Ein kritisches Hinterfragen bestimmter Wertepaare und weitergehende Recherchen bzgl. dieser Ausreißer können sicher bereits ab der ersten Stufe *reading the data* angestellt werden, zumal ein kritisches Hinterfragen der Kausalität bereits vor einer intensiven Auseinandersetzung mit den Daten erlaubt und angebracht ist (vgl. Kap. 2.2.6).

Die Fähigkeit *read behind the data* erstreckt sich daher in der hier vorgestellten grafischen Darstellung (vgl. Abb. 2-33) auch über alle drei Stufen und wird als *erweitertes* Modell abgebildet.

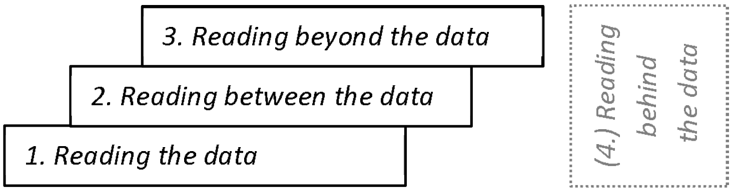


Abb. 2-33: Das erweiterte Modell in Bezug auf das Verständnis im Umgang mit Daten

Das Weiterfragen bzw. -recherchieren ist natürlich eng verbunden mit der bereits in Kapitel 2.1 erläuterten kritischen Grundeinstellung gegenüber den thematisierten Sachverhalten und den Daten selbst („*critical questions*“ und „*critical stance*“, S. 12): Fragen zur Erhebung bzw. zur Repräsentativität der Stichprobe folgen den kontextbezogenen Fragen beinahe von selbst, besonders wenn der dargestellte Sachverhalt dem erwarteten Zusammenhang des Betrachters zuwiderläuft (vgl. Abb. 2-34): „Warum nimmt in diesem Beispiel die Lautstärke sukzessive ab, obwohl die Schüleranzahl steigt?“ → „Ist das immer so?“ → „Betrifft das nur diese eine Schule?“ → „Wer hat diese Daten wann und wie erhoben?“ → „Sind die erhobenen Daten korrekt erfasst?“ usw.

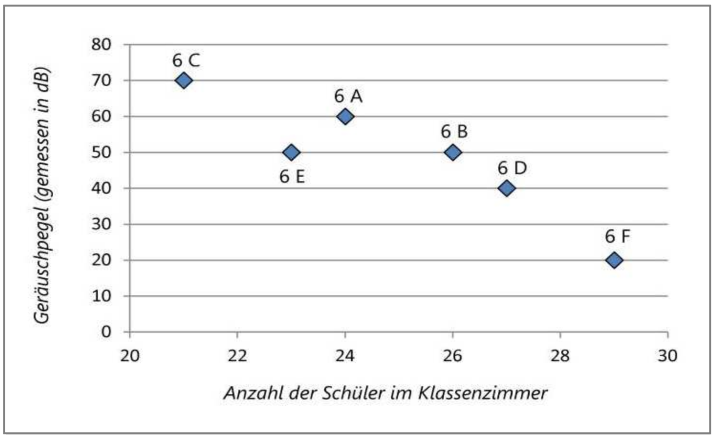


Abb. 2-34: Ein offensichtlich negativer Zusammenhang fordert kritisches Nachfragen

Wie bereits bei den Ausführungen zum Begriff einer (allgemeinen) Datenkompetenz sei auch an dieser Stelle darauf hingewiesen, dass sich in der ursprünglichen Diskussion um Grafikkompetenz und ihre drei Stufen nicht ausreichend darauf eingegangen wird, inwieweit es sich um die grafische Darstellung uni- oder bivariater Datensätze handelt. Eichler und Vogel konstatieren: Die Prinzipien des Lesens grafischer und tabellarischer Darstellungen gelten auch, „wenn bivariate Daten grafisch dargestellt werden“ (Eichler & Vogel 2013, S. 123). Allerdings werden beim Lesen und Interpretieren von Streudiagrammen zumindest auf den ersten beiden Stufen höhere Anforderungen gestellt (vgl. EICHLER & VOGEL 2013, S. 123):

➲ Beim Lesen eines Datenpunkts in einem Streudiagramm (*reading the data*) sind zwei Merkmalsausprägungen zu berücksichtigen, wohingegen beim Lesen in univariaten Daten die Häufigkeit eines bestimmten Merkmals direkt abgelesen werden kann.

➲ Während die Form einer Häufigkeitsverteilung, Muster und Streuung bei univariaten Daten nur entlang einer Richtung zu erfassen sind, gilt es bei Punktwolken die Streuung in zwei Richtungen zu beachten. Zudem ist die Form der zweidimensional streuenden Daten schwerer beschreibbar als bei univariaten Daten (*reading between the data*).

➲ Das Erkennen eines Musters in den Daten für eine „über die sichtbaren Daten hinausgehende Interpretation“ (*reading beyond the data*) ist mit der „Interpretation univariater Datensätze vergleichbar“ (EICHLER & VOGEL 2013, S. 123). Unter Umständen ist dies in der Punktwolke sogar einfacher sein als bei (unsortierten) univariaten Datensätzen.

Die Anforderungen an Schüler/innen im Umgang mit bivariaten Daten hinsichtlich ihrer Darstellungsform sind folglich höher anzusehen als im univariaten Fall.

Zusammenfassend stützt sich vor dem Hintergrund der gewählten Forschungsperspektive Datenkompetenz auf das in der Literatur etablierte Modell mit den drei hierarchischen Ebenen reading the data*,* reading between the data *und* reading beyond the data*. Dabei wird dieses Modell so erweitert bzw. modifiziert, so dass es den Umgang mit tabellarischen Daten ausdrücklich beinhaltet und das* reading *sich explizit auf bivariate Daten bezieht.*

Die erwähnte vierte Ebene ist, da sie im Zusammenhang mit CURCIOS dreistufigem Grundmodell oder dessen Varianten auch in der Literatur Erwähnung findet, der Vollständigkeit halber angeführt und auf eine neue Art zugeordnet, jedoch wird im weiteren Verlauf rein auf dieses dreistufige Grundmodell Bezug genommen. Dies hat zum einen praktikable Gründe, denn das Miteinbeziehen des kritischen

Hinterfragens des vorgelegten Datenmaterials würde den Rahmen der Studie sprengen, zum anderen ist diese vierte Ebene nicht empirisch belegt und erschwert bei der späteren Untersuchung hinsichtlich der Datenkompetenz bei einem gewissen Kompetenzscore die Zuordnung zu einer bestimmten Kompetenzstufe.

2.5 Datenkompetenz, bivariate Daten und Bildungspläne: Anforderungen an die Schülerinnen und Schüler

Die bisherigen Ausführungen haben verdeutlicht, dass eine sinnstiftende und verständige Bearbeitung bivariater Daten auch in der Sekundarstufe I möglich ist. Zu Recht stellt sich sodann die Frage, ob bzw. wie die Behandlung bivariater Daten im Unterricht durch die Lehrpläne legitimiert werden kann. Ist die Behandlung bivariater Daten überhaupt in den Lehrplänen bzw. Curricula der Länder verortet? Welche Voraussetzungen bringen Schüler/innen aus der Primarstufe mit?
Ein Blick in die Lehr- und Bildungspläne gebietet sich aber grundsätzlich auch dann, wenn für eine Studie – so wie im vorliegenden Fall – Daten in einem schulischen Rahmen erhoben werden und darüber hinaus – um in der statistischen Sprache zu bleiben – Schüler/innen die Merkmalsträger sind.

→ *Vor einem konkreten Blick in einige Lehrpläne werden zuerst die bezüglich Datenkompetenz und die für die vorliegende Studie besonders relevanten Punkte in den Bildungsstandards der KMK von 2003 dargestellt, auf deren Grundlage die jeweiligen Ländercurricula umgesetzt wurden; die zentrale Leitidee* Daten und Zufall *spielt dabei eine wesentliche Rolle.*

Da in der Sekundarstufe I „das (hoffentlich) aus der Primarstufe entwickelte Grundverständnis systematisiert und erweitert werden soll" (Eichler & Vogel 2013, S. X), wird vorab auf die Standards für den Primarbereich eingegangen, die nur ein Jahr nach den KMK-Beschlüssen für den Mittleren Schulabschluss vereinbart wurden und sich inhaltlich an den mathematischen Leitideen der weiterführenden Schulen orientieren. Dort wird unter den inhaltsbezogenen mathematischen Kompetenzen der Standard *Daten, Häufigkeit und Wahrscheinlichkeit* (KMK 2004,

S. 10) – hier beschränkt auf die für die vorliegende Arbeit relevanten Punkte – folgendermaßen konkretisiert: Die Schülerinnen und Schüler sollen

- in Beobachtungen, Untersuchungen und einfachen Experimenten Daten sammeln, strukturieren und in Tabellen, Schaubildern und Diagrammen darstellen,
- aus Tabellen, Schaubildern und Diagrammen Informationen entnehmen.

Da nicht nur für die empirische Untersuchung, sondern generell für den Aufbau von Datenkompetenz auch weitere Standards in engem Zusammenhang dazu stehen, sei hier zudem angeführt: Die Schülerinnen und Schüler sollen

- Gesetzmäßigkeiten in geometrischen und arithmetischen Mustern (z. B. in Zahlenfolgen oder strukturierten Aufgabenfolgen) erkennen, beschreiben und fortsetzen,
- funktionale Beziehungen in Sachsituationen erkennen, sprachlich beschreiben (z. B. Menge - Preis) und entsprechende Aufgaben lösen,
- funktionale Beziehungen in Tabellen darstellen und untersuchen

Diese Punkte aus dem Standard „Muster und Strukturen“ (KMK 2004, S. 10 f) muten durchaus anspruchsvoll an, zumal zu bedenken ist, dass es sich hier um verbindliche Vorgaben für die Grundschule handelt. Dennoch scheint die frühe Einbettung der Muster- und Strukturerkennung im Grundschulalter sinnvoll, denn ein Verständnis für mathematische Muster entwickelt sich nicht von selbst (vgl. GYSIN 2013, S. 126). Das Erkennen von Beziehungen und Struktur „ist kein trivialer Akt, sondern stellt eine hohe kognitive Fähigkeit (Lorenz, 2006 a) dar, die es zu entwickeln gilt.“ (GYSIN 2013, S. 126). Das „kann und soll auch schon in der Grundschule betrieben werden.“ (RADATZ, SCHIPPER, DRÖGE & EBELING 1999, zitiert nach GYSIN 2013, S. 126). BÖNIG und RUWISCH betonen ebenso, dass eine möglichst frühe Einbettung statistischer Denk- und Arbeitsweisen im Unterricht – vorzugsweise bereits in der Grundschule – erfolgen sollte, denn eine „kritische und reflektierte Haltung erwirbt man“ nicht automatisch „mit dem 18. Geburtstag, sondern man muss sie sich mühevoll aneignen.“ (BÖNIG & RUWISCH 2004, S. 6).

Für die weiterführenden Schulen ist die tragende Leitidee für den Aufbau statistischer Kompetenz selbstverständlich die bereits angesprochene Leitidee *Daten und Zufall*, die in dem Beschluss der KMK folgendermaßen formuliert ist (KMK 2003, S. 12): 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.

„Bereits die aufzählende Darstellung macht deutlich, wie stark innerhalb der Sekundarstufe I die Datenanalyse gegenüber der Wahrscheinlichkeitsrechnung akzentuiert wurde.“ (EICHLER & VOGEL 2013, S. XI).
Die oben dargestellte stichpunktartige Auflistung der inhaltsbezogenen Kompetenzen lassen sich nebenbei bemerkt mit sinngemäß beinahe identischen Formulierungen in WAGNERS Konkretisierungsversuchen von Datenkompetenz in elf Punkten (vgl. S. 8) wiederfinden. Da sie jedoch weit weniger ausdifferenziert sind als die in Kapitel 2.1 bereits dargestellten Eingrenzungsversuche von Datenkompetenz, finden sie dort keine separate Erwähnung.
Selbstverständlich lassen sich in den Bildungsstandards für die Sekundarstufe I weitere passende und notwendige inhaltsbezogene Kompetenzen für den Aufbau statistischer Kompetenz anführen:

- erkennen und beschreiben funktionale Zusammenhänge und stellen diese in sprachlicher, tabellarischer oder graphischer Form sowie gegebenenfalls als Term dar,

- analysieren, interpretieren und vergleichen unterschiedliche Darstellungen funktionaler Zusammenhänge (...)
- beschreiben Veränderungen von Größen mittels Funktionen, auch unter Verwendung eines Tabellenkalkulationsprogramms.

Speziell diese drei Aspekte der Leitidee *Funktionaler Zusammenhang* (KMK 2003, S. 11 f) differenzieren vor allem die beiden allgemeinen mathematischen (= prozessbezogenen) Kompetenzen *Argumentieren* und *Modellieren* (KMK 2003, S. 8) aus. Besonders das Modellieren halten EICHLER und VOGEL für „die vielleicht zentrale prozessbezogene Kompetenz [...] der Sekundarstufe I" (EICHLER & VOGEL 2013, S. XIII) im Umgang mit statistischen Daten.

KAUN (2006, S. 11 - 17) hat sich die Mühe gemacht, die Lehrpläne aller Bundesländer gezielt nach der Verankerung stochastischer und statistischer Themenbereiche zu durchsuchen und stellt fest, dass sich auf nationaler Ebene in allen Lehrplänen *aller* Schulformen die beschreibende Statistik verortet findet.
Bei der Betrachtung der bisherigen Zusammenstellung fällt auf: Die bivariate Datenanalyse ist „nicht unmittelbar Bestandteil der Leitidee *Daten und Zufall*" (EICHLER & VOGEL 2013, S. 119). So findet sich auch in den einzelnen Ländercurricula keine explizite Nennung von Bestandteilen einer bivariaten Datenanalyse. Einzig und allein in Niedersachsen enthält das aktuelle Kerncurriculum für die Realschule für das Ende der 10. Jahrgangsstufe den Passus: Die Schülerinnen und Schüler „stellen Datenpaare in zweidimensionalen Streudiagrammen dar und zeichnen die Ausgleichsgerade nach Augenmaß" (NIEDERSÄCHSISCHES KULTUSMINISTERIUM 2014, S. 30).
Eine so explizite Benennung ist nach Ansicht von EICHLER und VOGEL allerdings nicht zwangsläufig nötig, denn sie verstehen die allgemeine Formulierung „Daten" der Leitidee *Daten und Zufall* ohnehin so, „dass es nicht nur um Daten zu einem Merkmal (univariate Daten), sondern ebenso auch um Daten zu zwei (oder mehr) Merkmalen (bivariate bzw. multivariate Daten) gehen kann." (EICHLER & VOGEL 2013, S. XII)

Ein Blick in drei ausgewählte Bildungspläne soll die mögliche Übertragbarkeit der (aus Praktikabilitätsgründen in Bayern durchgeführte) Studie sowie die daraus resultierenden Ergebnisse auch auf andere Bundesländer aufzeigen. Drei Curricula werden dabei besonders in Augenschein genommen: Der Rahmenplan Mecklenburg-Vorpommerns (nichtgymnasialer Bildungsgang, 2011), das Kerncurriculum Niedersachsens (Realschule, 2014) sowie der bayerische LehrplanPLUS (Realschule, 2017). Der Entscheidung, just diese drei Bundesländer zu betrachten, liegen die folgenden Überlegungen zu Grunde:

a) Praktikabilität: als bayerische (Seminar-) Lehrkraft und Lehrbeauftragter der Universität kann der Verfasser dieser Studie sein Vorwissen über die bayerischen Curricula miteinbringen
b) Aktualität und direkter Bezug: Das Kerncurriculum für die Realschule Niedersachsens ist zum Zeitpunkt der Durchführung der Studie (abgesehen vom 2017 in Bayern neu eingeführten LehrplanPLUS) eines der aktuellsten Curricula und erwähnt darüber hinaus als einziges Curriculum Elemente der Datenanalyse bivariater Daten
c) Geografische Streuung: Die Rahmenpläne der integrierten Gesamtschule sowie des nichtgymnasialen Bildungsgangs Mecklenburg-Vorpommerns soll neben den beiden anderen Bundesländern auch die Themenbereiche eines zweigliedrigen Schulsystems sowie eines Bundeslandes im Nordosten repräsentieren.

Bei einem Vergleich der genannten drei Bildungspläne stellt man schnell fest, dass sich die Kompetenzerwartungen bzw. Lerninhalte zu den Themenbereichen *Lineare Funktionen* sowie *Statistische Kenngrößen* in fast gleichlautenden Formulierungen in allen drei Curricula finden lassen. In der tabellarischen Übersicht der Abb. 2-35 (S. 64) sind die entsprechenden Passagen wortwörtlich zitiert.

Bundesland / Themenbereich	Mecklenburg-Vorpommern [1)]	Niedersachsen [2)]	Bayern [3)]
	Die Schülerinnen und Schüler …		
Lineare Funktionen	erkennen funktionale Zusammenhänge und können Sie in geeigneter Weise beschreiben; Funktionen als Modelle realer Sachverhalte *(Jgst. 8)*	stellen lineare Zusammenhänge als Funktionsgleichung dar und im Koordinatensystem dar *(Ende Jgst. 8)*	erfassen und beschreiben funktionale Zusammenhänge *(Jgst. 8)*; zeichnen Graphen von linearen Funktionen *(Jgst. 9)*
Statistische Kenngrößen	können über Kenngrößen arithmetisches Mittel, Zentralwert, Modalwert, Spannweite, Quartile die Daten statistisch aufbereiten und kritisch auswerten. *(Ende Jgst. 6)*	beschreiben die Datenverteilung mit den Begriffen Minimum, Maximum, Spannweite, Ausreißer, Median; berechnen das arithmetische Mittel *(Ende Jgst. 6)*	ermitteln bei Daten die statistischen Kenngrößen arithmetisches Mittel, Zentralwert, Modalwert und Spannweite. *(Jgst. 7)*

1) Ministerium für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern (2010), S. 16 - 22
2) Niedersächsisches Kultusministerium (2014), S. 29 f
3) Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2017), S. 750 f

Abb. 2-35: Auszüge aus den Ländercurricula Mecklenburg-Vorpommerns, Niedersachsens und Bayerns

Erste Unterschiede zeigen sich, wenn man die Lerninhalte bzw. Kompetenzerwartung bzgl. der Darstellung von Daten vergleicht: Während Mecklenburg-Vorpommern und Niedersachsen in ihren Curricula relativ konkret die Darstellungsformen benennen (z. B. „… stellen Daten in Tabellen, Balkendiagrammen und Säulendiagrammen dar“, „Boxplots“), findet sich im bayerischen LehrplanPLUS die eher allgemein gehaltene Formulierung „[Die Schülerinnen und Schüler] analysieren Daten kritisch, um Fehler in Diagrammen zu erkennen und zu korrigieren.“ (Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst 2017, S. 753).

Besonders bemerkenswert erscheinen folgende Unterschiede:

- In Bayern ist der Themenbereich Statistik offenbar mit dem Ende der 7. Jahrgangsstufe abgeschlossen. Zwar erstreckt sich der Lernbereich „Daten und Zufall“ bis zur 10. Jahrgangsstufe, allerdings werden ab der 8. Klasse unter diesem Punkt ausschließlich stochastische Themen (Zufallsexperimente, Gewinnwahscheinlichkeiten, Pfadregeln, …) behandelt.
- Niedersachsen formuliert in seinem Kerncurriculum – ebenfalls in der Rubrik Daten und Zufall – als einziges Bundesland explizit „Kernkompetenzen“ aus dem Bereich der bivariaten Datenanalyse: am Ende der 10. Jahrgangsstufe sollen die Schülerinnen und Schüler „Datenpaare in zweidimensionalen Streudiagrammen“ darstellen und „die Ausgleichsgerade nach Augenmaß“ (NIEDERSÄCHSISCHES KULTUSMINISTERIUM 2014, S. 30) einzeichnen können.
- Mecklenburg-Vorpommern räumt im Rahmenplan Mathematik des nichtgymnasiale Bildungsgangs durch die ausdrückliche Nennung der Themengebiete „Statische Erhebungen“ der Datenanalyse (zusätzlich zu den Themengebieten der Wahrscheinlichkeit!) einen besonderen Stellenwert ein. Zudem werden deren Inhalte relativ genau konkretisiert und „die Verwendung geeigneter Hilfsmittel wie Software“ explizit gefordert (MINISTERIUM FÜR BILDUNG, WISSENSCHAFT UND KULTUR MECKLENBURG-VORPOMMERN 2010, S. 20)

Trotz der teilweisen Heterogenität der Lehrpläne sind die fachlichen Voraussetzungen für Lesen und Interpretieren bivariater Datensätze (vgl. die *Arbeitsdefinition* in Kap. 2.4, S. 58) am Ende einer 6. Jahrgangsstufe in jedem Bundesland gegeben. Auch die elementaren Kenntnisse über lineare Funktionen (Aufstellen einer linearen Geradengleichung) kann am Ende der 9. Jahrgangsstufe in jedem Bundesland als vorausgesetzt angenommen werden.

2.6 Forschungsstand

→ *Da deutschsprachige Studien zum Umgang mit bivariaten Daten kaum existent sind, werden im Folgenden vorrangig die Befunde englischsprachiger Studien vorgestellt, die in ihrer Summe doch alle ein relativ ähnliches Bild zeichnen: Beim Lesen von und in Daten (*reading the data *und* reading between the data*) zeigen sich so gut wie alle Schüler/innen, auch die jüngeren, sehr datenkompetent; beim Lesen über den bestehenden Datensatz hinaus (*reading beyond the data*) sinkt dieses Kompetenzlevel signifikant. Grundsätzliche Überzeugungen bzw. Erwartungen, das Alter der Probanden sowie die Art und Stärke des Korrelation haben einen entscheidenden Einfluss auf die Prognosekompetenz der Schüler/innen.*

Der Umgang von Schüler/inne/n mit statistischen Daten hat vor allem im englischsprachigen Raum „eine längere und stärkere Tradition als in Deutschland" (EICHLER & VOGEL 2013, S. 68), dadurch ist zumindest die grafische Aufbereitung von Daten und deren Untersuchung anhand der klassischen statistischen Kenngrößen (Lage- und Streuparameter) „international breit untersucht worden (Shaughnessy 2007)" (EICHLER & VOGEL 2013, S. 68). Auf nationaler Ebene dagegen lassen sich dazu so gut wie keine Forschungsergebnisse finden, denn die Datenanalyse spielt im Mathematikcurriculum „immer noch eine eher untergeordnete Rolle" (EICHLER & VOGEL 2013, S. 70; vgl. dazu auch Kap. 2.5).

Noch schwieriger gestaltet sich die Suche nach deutschsprachigen Studien über den Umgang von Schüler/inne/n mit *bivariaten* statistischen Daten. Auch EICHLER und VOGEL betonen in der aktuellsten Auflage ihres Buches „Leitidee Daten und Zufall", dass es quasi keine Forschungsergebnisse zu diesem Thema gibt (vgl. EICHLER & VOGEL 2013, S. 130). Gleichwohl es mittlerweile vor allem von ebendiesen beiden als auch weiteren einschlägigen Autoren (BIEHLER, SCHWEYNOCH, BOROVCNIK, VOGEL, WINTERMANTEL, ...) didaktisch wertvolle und gewinnbringende (deutschsprachige) Materialien gibt, mit deren Hilfe eine spannende Hinführung und Behandlung dieser Thematik im Unterricht gelingen kann, sind veröffentlichte Studien, die sich mit der Analyse der Fähigkeiten der Schüler/innen, Zusammenhänge in bivariaten Datensätzen zu erfassen, grafisch darzustellen und/oder zu analysieren, quasi nicht existent.

Immerhin lassen sich zumindest auf internationaler Ebene (hauptsächlich in den USA, Australien und Neuseeland) einige Untersuchungen finden, die sich dem Umgang von Schüler/inne/n mit bivariaten Datensätzen und deren grafischen Darstellungen widmen.

Die Ergebnisse einer der ersten großen Untersuchungen zu diesem Thema veröffentlichten 1997 Watson und Moritz, die in den einführenden Worten ihrer Arbeit ebenfalls noch erwähnen, dass sich zum damaligen Zeitpunkt der Durchführung dieser Studie keine Forschungsliteratur finden lasse, die sich mit dem Verständnis der Schüler/innen bezüglich der Beziehung von Variablen und deren grafische Veranschaulichung beschäftigt: „*Research into the [...] understanding of association among variables, the ability to graph such associations [...] is not found in the statistics education literature.*“ (Watson & Moritz 1997, p. 133).

Um einschätzen zu können, inwieweit Schüler/innen eine vorgegebene (Schein-) Korrelation im Text erfassen und diese kritisch hinterfragen, wurde ihnen der in Abb. 2-36 dargestellte Zeitungsartikel an die Hand gegeben mit der Bitte, den im Text beschriebenen „beinahe perfekten Zusammenhang“ zwischen dem Anstieg der Todesfälle durch Herzversagen und dem entsprechenden Ansteigen

Family car is killing us, says Tasmanian researcher

Twenty years of research has convinced Mr Robinson that motoring is a health hazard. Mr Robinson has graphs which show quite dramatically an almost perfect relationship between the increase in heart deaths and the increase in use of motor vehicles. Similar relationships are shown to exist between lung cancer, leukaemia, stroke and diabetes.

Draw and label a sketch of what one of Mr. Robinson's graphs might look like.
What questions would you ask about his research?

Abb. 2-36: Die gestellte Aufgabe mit dem „beinahe perfekten Zusammenhang“ (Watson & Moritz 1997, p. 136) der Benutzung von Kraftfahrzeugen als Erstes grafisch darzustellen und im Anschluss zu äußern, welche Fragen sie an den Autor dieses Artikels stellen würden. Die Antworten bzw. grafischen Veranschaulichungen zum ersten Teil der Aufgabenstellung teilten WATSON und MORITZ anschließend in fünf Kategorien ein (vgl. WATSON & MORITZ 1997, p. 137), die vom Nichterfassen der Aufgabenstellung (Kategorie 1) bis zur vollständig korrekten grafischen Darstellung inklusive einer möglichen Achsenbeschriftung und Skalierung (Kategorie 5) reichen.

Obwohl WATSON und MORITZ in ihrer Veröffentlichung auf konkrete Zahlen, wie viele der untersuchten 1291 Sechst- bis Elftklässler aus Australien und England welchem der fünf Kompetenzkategorien zugeordnet werden können, verzichten, halten sie fest, dass zumindest viele der Probanden sich in der Kategorie 3 bis 5 wiederfinden und somit mindestens die Stufe 2 ihrer dreistufigen statistischen Kompetenzskala aufweisen („*the second tier of statistical understanding*", WATSON & MORITZ 1997, p. 142): Sie erfassen im Großen und Ganzen zumindest die Kernaussage des Zeitungsartikels.

Weit interessierter waren WATSON und MORITZ allerdings daran, ob Schüler/innen mit dem zweiten Teil der Aufgabenstellung („Welche Fragen würdest du an den Autor dieses Artikels stellen?") die „höchste Stufe statistischen Denkens" (= Kompetenzstufe 3), erreichen: Stellen Schüler/innen den beschriebenen Zusammenhang in dieser Aufgabe überhaupt in Frage? Auch hier teilten sie die Antworten wieder in fünf Kategorien ein, die von einer unreflektierten, „blinden Akzeptanz" des Zusammenhangs (Kategorie 1) bis zum kritischen Hinterfragen inklusive einer Nennung weiterer Ursachenfaktoren (Kategorie 5) reichen. Die erwähnte höchste Kompetenzstufe 3 wurde nur Schüler/inne/n zugestanden, deren Antworten sich den letzten beiden Kategorien 4 und 5 zuordnen ließen. Auch hier haben WATSON und MORITZ keine konkreten Zahlen veröffentlicht, wie viele Antworten welcher Kategorie zufielen, halten jedoch fest, dass die meisten Studienteilnehmer die behauptete Korrelation ohne Nachfragen bereitwillig akzeptierten („Das scheint zu stimmen."). Darüber hinaus betonten sie, dass für die erfolgreiche Bewältigung des zweiten Teils der Aufgabe unverkennbar bereits das Erreichen einer höheren Kategorie im ersten Teil der Aufgabe eine offenbar entscheidende Voraussetzung sei. Zudem sei eine „angemessene statistische Sprache" sehr selten bei den

Schüler/inne/n anzutreffen: *„Sophisticated statistical language often will not used by the students."* (WATSON & MORITZ 1997, p. 144)

→ *Obwohl das Hinterfragen einer entdeckten Korrelation in einem bivariaten Datensatz kein expliziter Bestandteil der vorliegenden Forschungsarbeit ist, ist die Studie von WATSON und MORITZ insofern interessant, weil sich auch in weiteren Studien belegen lässt (u. a. MORITZ 2004, s. u.), dass bereits vorhandene Überzeugungen, Vorstellungen und Erfahrungen einen bedeutenden Einfluss auf die Beurteilung möglicher Zusammenhänge zweier Variablen hat:* „A robust finding is that peoples' prior beliefs about relationship between two variables have a great deal of influence on their judgments of covariation between those variables (e. g. Jennings, Amabile & Ross, 1982; Kuhn, Amsel & O' Loughlin, 1988)." *(ZIEFFLER 2008, p. 293).*

MORITZ vertiefte die Forschungen in diese Richtung und veröffentlichte 2004 die Ergebnisse einer weiteren Studie, in der er 167 australische Schüler/innen aus den Jahrgangsstufen 3, 5, 7 und 9 auf ihre Argumentationsfähigkeit bezüglich „Kovariation" (d. h. bzgl. der (Mit-) Abhängigkeit einer Variablen von einer anderen) hin untersuchte. Konkret interessierten ihn dabei drei Bereiche:

(1) Können Schüler/innen die Aussage eines Textes in einem Diagramm darstellen (*speculative data generation*)?
(2) Sind sie fähig, ein Streudiagramm zu verbalisieren (*verbal graph interpretation*)?
(3) Können sie einem Diagramm Werte entnehmen und neue Daten einordnen (*numerical graph interpretation*)?

Die grafischen Darstellungen bzw. Antworten zu jedem dieser Bereich teilte MORITZ zur Auswertung anschließend in vier Level ein (*Nonstatistical*, *Single Aspect*, *Inadequate Covariation*, *Appropriate Covariation;* vgl. Abb. 2-38 auf S. 71).

(1) Speculative data generation

Ähnlich wie in der gemeinsamen Arbeit mit Watson (s. o.) ließ Moritz in dieser Studie in einem ersten Schritt wieder das Textverständnis (mit-) abprüfen: Bei der ersten Aufgabe 1 („*Task 1 (Negative association)*“, Abb. 2-37) sollten die Schüler/inne/n die Aussage „Leute, die mehr Zeit ins Lernen investieren, erhalten tendenziell eine schlechtere Note.“ grafisch darstellen.

Task 1 (Negative association)

Anna and Cara were doing a project on study habits.
They asked some students two questions:

- “What time did you spend studying for the spelling test?”
- “What score did you get on the test?”

Anna asked 6 students. She used the numbers to draw a graph.
She said, “People who studied for more time got lower scores.”

Q1. Draw a graph to show what Anna is saying for her 6 students. Label the graph.

Task 1 (Positive association)

She said, “People who studied for more time got higher scores.”

Q1*. Draw a graph to show what Anna is saying for her 6 students. Label the graph.

Task 1 to assess speculative data generation.
(Third- and fifth-grade males received Q1* in place of Q1.)

Abb. 2-37: Aufgabe 1 der Studie von Moritz (2004) mit beiden Varianten (negative und positive Korrelation)

Vor allem viele ältere Schüler/inne/n schafften hier das höchste Level „*Appropriate Covariation*“: Sie erfassten korrekt den im Text behaupteten indirekten Zusammenhang zwischen Lernzeit und Abschneiden im Rechtschreibtest und konnten diesen zutreffend in einem Diagramm darstellen (detaillierte Ergebnisse: siehe Abb. 2-38, S. 71). Einige Schülern (die männlichen Dritt- und Fünftklässler) ließ Moritz übrigens dieselbe Aufgabe, allerdings mit einem positiven Zusammenhang (mehr Lernzeit → bessere Note) bearbeiten, die im Vergleich zum Rest deutlich besser abschnitten (Abb. 2-38, Spalte 3^a und 5^a).

Moritz begründet dies damit, dass diese Art einer Beziehung mit den grundsätzlichen Überzeugungen und Erwartungen der Jugendlichen eher konform läuft als ein indirekter Zusammenhang bei diesem Kontext bzw. ihnen diese Korrelation logischer erscheint (weil sie so einen Zusammenhang ohnehin vermuten) (vgl. Moritz 2004, p. 244). Beachtliche 78 % der männlichen Fünftklässler erreichten hier bereits ebenfalls das vierte und höchste Level „*Appropriate Covariation*", im Gegensatz zu den Mädchen (die in ihrer Aufgabe die negative Korrelation vorliegen hatten) mit 38 %.

Percentage of student responses at four levels of speculative data generation by gender and by grade (N = 167)

Levels of Speculative Data Generation	Female Grade 3	Female Grade 5	Female Grade 7	Female Grade 9	Male Grade 3[a]	Male Grade 5[a]	Male Grade 7	Male Grade 9	Total (*N*)
0–Nonstatistical	23	12	5	40	11	11	8	0	18
1–Single Aspect	0	12	14	20	42	11	4	0	18
2–Inadequate Covariation	42	38	18	20	5	0	21	11	35
3–Appropriate Covariation	35	38	64	20	42	78	67	89	96
Total (*N*)	26	26	22	5	19	18	24	27	167[b]

[a] Third- and fifth-grade males were administered Q1* rather than Q1.
[b] Percentages do not always sum to 100 due to rounding.

Abb. 2-38: Detaillierte Ergebnisse der „*Speculative Data Generation*"

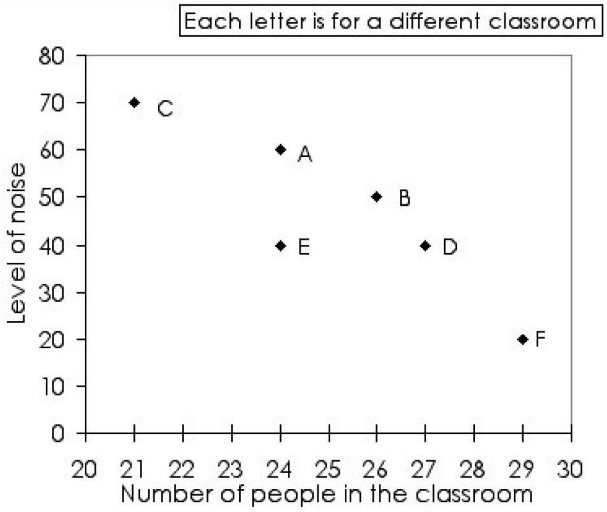
(2) Verbal Graph Interpretation

Auch bei der Interpretation des Streudiagramms bzw. beim *Übersetzen* des Diagramms in Worte lohnt ein detaillierter Blick auf die Schülerantworten. Die Probanden sollten sich vorstellen, sie müssten den Inhalt bzw. die Aussage der Grafik jemandem beschreiben, der sie momentan nicht sehen kann. Im Anschluss daran sollten sie entscheiden, ob die Grafik eine gute Argumentationsgrundlage darstelle für die verallgemeinernde Aussage, dass Klassenzimmer mit mehr Schüler/inne/n weniger Lärm verursachen (vgl. „*Task 2 Q2d*", Abb. 2-39, S. 72).

Auch hier schneiden tendenziell die älteren Schüler/innen deutlich besser ab als die jüngeren (vgl. Abb. 2-40, S. 72), und männliche Teilnehmer besser als weibliche, wenngleich MORITZ dafür eher die Auswahl der untersuchten Schulklasse mehr verantwortlich macht als das Geschlecht: „*Seventh- and ninth-grade males performed better than their female counterparts, although this is likely due to classes sampled rather the students' gender.*" (MORITZ 2004, p. 245).

Vor allem die Dritt- und Fünftklässler kommen hier über die ersten beiden Level nicht hinaus: Sie liefern Antworten wie „Die Grafik zeigt die Klassen A, B, C, D, F und E und Zahlen." (≙ Level 0 (*nonstatistical*)) oder beziehen sich auf nur eine

Task 2
Some students were doing a project on noise.
They visited 6 different classrooms.
They measured the level of noise in the class with a sound meter.
They counted the number of people in the class.
They used the numbers to draw this graph.



Q2a. Pretend you are talking to someone who cannot see the graph. Write a sentence to tell them what the graph shows. "The graph shows...

Q2b. How many people are in Class D?

Q2c. If the students went to another class with 23 people, how much noise do you think they would measure? (Even if you are not sure, please estimate or guess.) Please explain your answer.

Q2d. Jill said, "The graph shows that classrooms with more people make less noise". Do you think the graph is a good reason to say this?
❑ YES or ❑ NO Please explain your answer.

Q2d*. Jill said, "The graph shows that the level of noise is related to the number of people in the class". Do you think the graph is a good reason to say this?
❑ YES or ❑ NO Please explain your answer.

Task 2 to assess verbal and numerical graph interpretation.
(Third- and fifth-grade males received Q2d* in place of Q2d.)

Abb. 2-39:Aufgabe zur Verbalisierung eines grafisch dargestellten Sachverhalts

Percentage of student responses at four levels of verbal graph interpretation by gender and by grade (N = 121)

Levels of Verbal Graph Interpretation	Female Grade 3	5	7	9	Male Grade 3[a]	5[a]	7	9	Total (*N*)
0–Nonstatistical	31	13	0	0	31	29	0	0	13
1–Single Aspect	38	22	20	8	46	14	8	10	25
2–Inadequate Covariation	23	57	45	67	15	43	33	5	43
3–Appropriate Covariation	8	9	35	25	8	14	58	86	40
Total (*N*)	13	23	20	12	13	7	12	21	121[b]

[a] Third- and fifth-grade males were administered Q2d* rather than Q2d.
[b] Percentages do not always sum to 100 due to rounding.

Abb. 2-40: Detaillierte Ergebnisse zur Verbalisierung des Diagramms (*Task 2*)

Variable und lassen die zweite außer Acht (≙ Level 1 (*single aspect*)): „Das Diagramm zeigt, dass einige Klassen leiser sind.“ oder „80 ist das Lauteste und 0 das Geringste.“ Antworten, die sich rein auf einen Punkt beziehen („Die Darstellung zeigt, dass Klasse C 21 Kinder hat und einen Geräuschpegel von 70.“) fallen ebenfalls noch in Level 1. Ein Großteil der (Fünftklässler 57 % der Mädchen und 43 % der Jungen) erreichen immerhin Level 2 (*inadequate covariation*): Antworten auf dieser Stufe beziehen sich zwar auf beide Variablen, stellen aber keinen Bezug zwischen ihnen her: „Es wird die Anzahl der Kinder in einer Klasse und der Geräuschpegel dargestellt.“ Verallgemeinernde Aussagen wie „Das Klassenzimmer mit den wenigsten Kindern ist das lauteste und das mit den meisten das leiseste.“ oder „Die Klasse mit den wenigsten Kindern macht am meisten Lärm.“ stufte MORITZ ebenfalls in dieses Level ein.

Vor allem die Siebt- und Neuntklässler sind es, die Level 3 (*appropriate covariation*) erreichen. Auf dieser Stufe gelingt das zutreffende Beschreiben einer gewissen Abhängigkeit beider Variablen. Manche Schülerstatements sind zwar einfach gehalten („Je weniger Leute, desto mehr Lärm“), aber korrekt, andere wiederum schildern erst Details und enden mit einem Fazit („Zimmer C ist das lauteste, dann Zimmer A, gefolgt von B; E und D mit jeweils 40, und F bildet das Schlusslicht, also gilt: Mehr Menschen, weniger Lärm.“). Im Gegensatz zu Level 2-Antworten, bei dem einige Schüler/inne/n bereits einen Randpunkt der Grafik für eine Pauschalantwort nutzten, beinhalten Level 3-Antworten die Betrachtung beider Randpunkte und verwenden den Plural, um eine Tendenz festzuhalten: „Die Klassen mit den wenigsten Schülern sind die lautesten. Die Räume, in denen sich die meisten Kinder aufhalten, sind die leisesten.“

(3) Numerical Graph Interpretation

Als besonders interessant für die vorliegende Forschungsarbeit erweisen sich die Ergebnisse aus dieser Rubrik, inwieweit Schüler/innen hypothetische Werte in die vorhandene Grafik einordnen können: „Wenn die Schüler in eine andere Klasse mit 23 Kindern gehen würden, welchen Geräuschpegel würden sie dann wohl messen?“ (Frage *Q2c*, vgl. Abb. 2-39 auf S. 72). Mit Blick auf die Tabelle aus Abb. 2-41 (S. 74) scheint bemerkenswert, dass nicht ein einziger Proband aus Jahrgangsstufe 3 und 5 das dritte Level erreichte. Es scheint sich also auch bei der Auswertung von *Task 2* zu bestätigen, dass ältere Schüler/innen bei diesem Aspekt statistischer Bildung deutlich besser abschneiden.

Percentage of student responses at four levels of numerical graph interpretation by gender and by grade (N = 121)

Levels of Numerical Graph Interpretation	Female Grade 3	Female Grade 5	Female Grade 7	Female Grade 9	Male Grade 3	Male Grade 5	Male Grade 7	Male Grade 9	Total (*N*)
0 – Nonstatistical	46	13	0	0	15	14	0	0	12
1 – Single Aspect	23	52	25	33	62	29	25	10	39
2 – Inadequate Covariation	31	35	35	50	23	57	50	10	40
3 – Appropriate Covariation	0	0	40	17	0	0	25	81	30
Total (*N*)	13	23	20	12	13	7	12	21	121[a]

[a] Percentages do not always sum to 100 due to rounding.

Abb. 2-41: Ergebnisse bzgl. des Einordnens hypothetischer neuer Werte in ein bestehendes Datengefüge.

Immerhin schafften gut ein Drittel der Mädchen aus diesen Klassen das Level 2. MORITZ stufte auf diesem Level diejenigen Schüler/innen ein, die Pegelwerte zwischen 39 und 54 nannten und zur Begründung die beiden nächst gelegenen Punkte heranzogen, wobei die meisten zum Wert 50 als Wert in der Mitte tendierten: „Ich denke 50, denn in den beiden Klassen mit 24 Schülern sind es 40 und 60, und 50 liegt in der Mitte." Sie bezogen nicht den Trend im Allgemeinen mit ein, wie beispielsweise Schüler/innen aus Level 3 (*approriate covariation*) es taten. Für das Erreichen dieses Levels waren in den Antworten Werte von 55 bis 70 nötig, wobei auch hier 65 eine häufige Antwort war: „Ungefähr 65, weil in den Klassen mit 24 Schülern ist es 60 und in der Klasse mit 21 ist es 70."

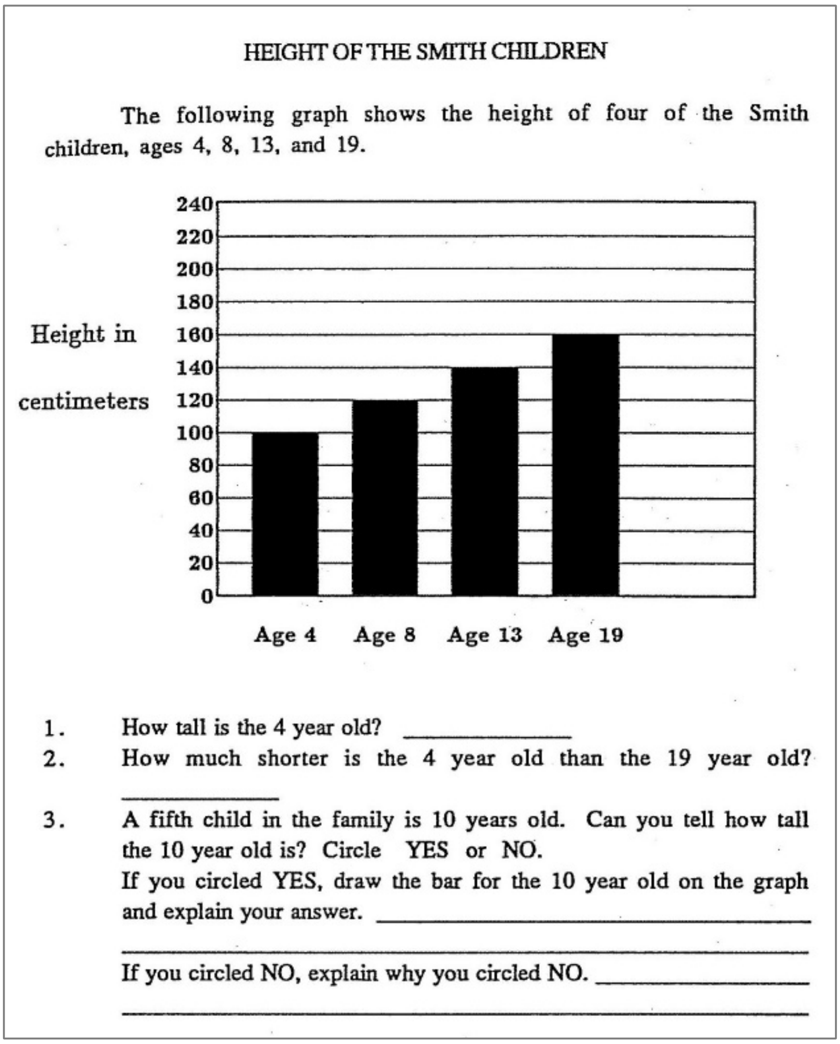
→ *MORITZ' Studie unterstreicht die Bedeutung des Alters der Probanden sowie die eindeutige Tendenz, dass eine positive Korrelation deutlich leichter erkannt wird als eine negative. Allerdings scheint für die „Erkennungsrate" einer Korrelation auch deren Güte entscheidend zu sein: In der Studie von ESTEPA und BATANERO (vgl. S. 78 ff) wurde die negative Korrelation sogar geringfügig öfter von den Probanden erfasst als die direkte, da im Streudiagramm ein deutlich höherer Korrelationskoeffizient ersichtlich war. Interessant wäre vor allem bei einer* prediction, *inwieweit die Erwartungshaltung* *bzw. die Vorerfahrungen der Schüler/innen hier eine (erfolgreiche) Einordnung der Werte hemmen, da davon auszugehen ist, dass eine gewisse Anzahl an Schüler/innen sicher erwarten, dass in Klassen mit weniger Schüler/inne/n tendenziell deutlich weniger Lärm herrscht als in großen Klassen.*

Über das schwache Abschneiden bei Prognosen aufgrund vorliegenden Datenmaterials weist auch PEREIRA-MENDOZA in seiner Untersuchung über die Rolle des graphischen Darstellens bei der Datenauswertung bei Grundschülern hin (die sich allerdings auf univariate Datensätze bezieht): „Wenn man z. B. Schüler bittet, eine Graphik zu interpretieren, wie Schüler zur Schule kommen, gäbe es nur wenig Schwierigkeiten. Wenn man sie aber fragte, wie ein neuer Schüler vielleicht zur Schule kommt, würden sie die Informationen der Darstellung ignorieren und ihr eigenes Wissen nutzen. [...] Die Schüler können zwar graphische Darstellungen interpretieren, aber sie können diese nicht nutzen, um eine realistische Vorhersage zu treffen oder eine realistische Folgerung zu äußern.“ (PEREIRA-MENDOZA 1995, S. 11).

Auch bei einer anderen Aufgabe, bei der die Schüler/innen erst die Anzahl farbiger Bonbons erfolgreich in einem Diagramm darstellten, fingen sie an zu raten, sobald es die mögliche Anzahl roter Bonbons in einer weiteren Packung zu *erahnen* galt, selbst „als sie auf die Darstellung hingewiesen wurden.“ (PEREIRA-MENDOZA 1995, S. 11).

Diese Aussagen ergründete PEREIRA-MENDOZA zusammen mit MELLOR in einer ausführlichen Studie, in der sie 121 Schüler/innen aus der 4. Jahrgangsstufe und 127 Schüler/innen aus der 6. Jahrgangsstufe in einem schriftlichen Test mit zwölf Aufgaben bearbeiten ließen, anhand dessen Ergebnisse die Fähigkeit der Kinder untersucht werden sollte, inwieweit sie Information aus einem Diagramm lesen, verbalisieren bzw. interpretieren und für Vorhersagen nutzen können (vgl. das dreistufige Kompetenzmodell nach CURCIO, FRIEL & BRIGHT!). Während das *Lesen* grafischer Darstellungen fast allen Kindern keine Probleme bereitete (95 % korrekte Antworten bei den Viertklässlern, 98 % bei den Sechstklässlern), zeigten sich deutlich mehr Schwierigkeiten bei Fragstellungen, die auf einen Vergleich der Daten abzielen. Hier sank die Erfolgsquote auf 52 % (Viertklässler) bzw. 68 %

(Sechstklässler); allerdings waren die meisten Falschantworten auf reine Rechen- oder Ablesefehler zurückzuführen. Die bereinigte Quote *echter* Falschantworten (deren Ursache also nicht ersichtlich war) lag so nur bei 8 % bzw. 4 % (vgl. PEREIRA-MENDOZA & MELLOR 1990, p. 152).



HEIGHT OF THE SMITH CHILDREN

The following graph shows the height of four of the Smith children, ages 4, 8, 13, and 19.

1. How tall is the 4 year old? ____________
2. How much shorter is the 4 year old than the 19 year old? ____________
3. A fifth child in the family is 10 years old. Can you tell how tall the 10 year old is? Circle YES or NO.
   If you circled YES, draw the bar for the 10 year old on the graph and explain your answer. ____________
   If you circled NO, explain why you circled NO. ____________

Abb. 2-42: Beispielaufgabe aus dem Testset von PEREIRA-MENDOZA und MELLOR (1990)

Die größten Schwierigkeiten hatten die Probanden bei den Vorhersageaufgaben (vgl. Abb. 2-42 Aufgabe 3), bei denen die Quote korrekter Antworten nochmals signifikant abfiel: „*... the level of performance was extremely low; the mean success rates for grade four students was 16 %, and for grade six students was 18 %.*“ (PEREIRA-MENDOZA & MELLOR 1990, p. 154).

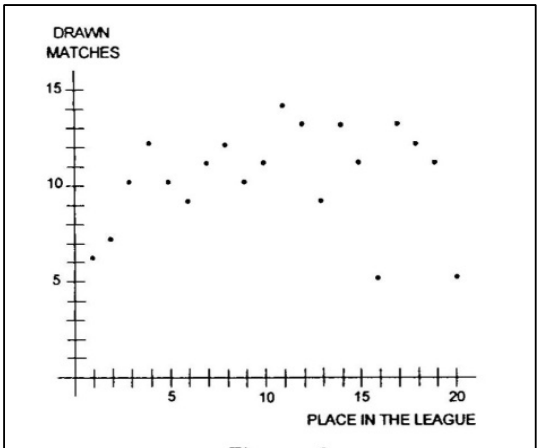
Die beiden Autoren konnten für dieses schwache Abschneiden hauptsächlich folgende fehlerhafte Strategien als Gründe eruieren:

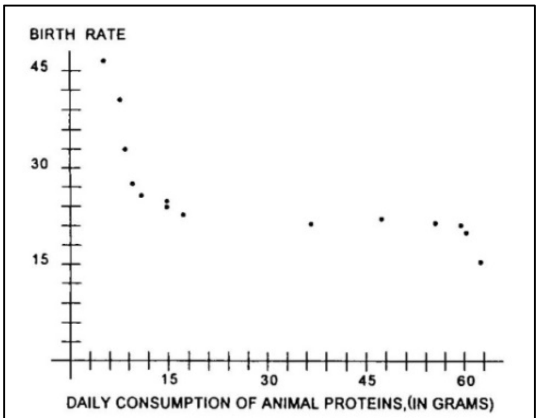
- Schüler/innen suchen beinahe zwanghaft nach Mustern und Strukturen in Diagrammen, auch wenn keine vorhanden sind („*Students tended to think that patterns* must *exist in a graph.*“, PEREIRA-MENDOZA & MELLOR 1990, p. 155).
- Sie beharren auf Ihrer Prognose, auch wenn diese nicht realistisch oder begründbar erscheint (dass z. B. zehnjährige Schüler/innen in der Regel größer sind als 19-Jährige).
- Sehr häufig geben Schüler/innen an, dass eine Einschätzung bzw. die Einordnung eines angenommenen Wertes nicht möglich sei, weil eben genau dieser in der Grafik fehle: „*... they said thy could not five an answer since ‚it' was not on the graph*“ (PEREIRA-MENDOZA & MELLOR 1990, p. 155).

→ *Vor allem der letzte Punkt dieser Aufzählung scheint für die geplante Untersuchung besonders interessant: Zum einen fehlen im Testheft bei den grafischen und tabellarischen Darstellungen ganz bewusst einige Wertepaare, die zu Aussagen à la „Das kann man dem Diagramm nicht entnehmen.“ führen könnten. Ob sich darüber hinaus Schüler/innen, wie hier im ersten Aufzählungspunkt geschildert, durch die spezielle Unterrichtssituation bzw. Testsituation auch dazu* genötigt *fühlen, einen Trend in den Daten auszumachen, bleibt abzuwarten.*

SHAUGHNESSY (2007, p. 957 - 1009) bestätigt in seiner ausführlichen Zusammenschau *Research on Statistics Learning and Reasoning*, dass sehr viele Jugendliche Tabellen und Diagramme zwar *auslesen* können, aber erhebliche Schwierigkeiten haben, die entnommenen Informationen für z. B. Schlussfolgerungen zu verwerten (vgl. SHAUGHNESSY 2007, p. 959).

Entgegen diesen Ausführungen von SHAUGHNESSY und PEREIRA-MENDOZA (Schüler/innen können zwar graphische Darstellungen interpretieren, aber nicht nutzen, um Vorhersagen zu treffen), meinen SCHWIRTZ und BEGENAT, dass die „Frage nach einem statistischen Zusammenhang zwischen zwei Größenarten [...] von Kindern der Klasse 3 verständig bearbeitet werden kann." (SCHWIRTZ & BEGENAT 2000, S. 50). Obwohl diese Aussage lediglich eine Randbemerkung eines Artikels in einer Fachzeitschrift darstellt und ihr keine empirische Untersuchung zugrunde liegt, ist sie für die vorliegende Forschungsarbeit durchaus erwähnenswert, da die Fragebögen in erster Linie von deutlich älteren Schüler/inne/n aus der 6. und 9. Jahrgangsstufe bearbeitet werden.





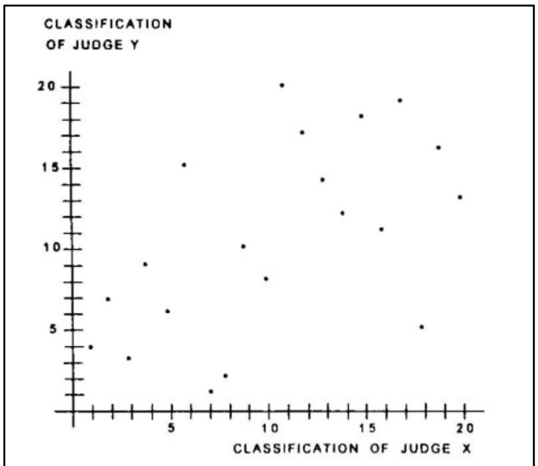


Abb. 2-43: *Independence, inverse, direct Correlation*

Auf den Umstand, dass nicht nur Jugendliche, sondern auch Erwachsene dazu tendieren, sich bei der Einschätzung, ob zwei Größen in Abhängigkeit zueinander stehen, eher auf frühere Überzeugungen und Erwartungen stützen als auf die vorliegenden Daten, weisen auch ESTEPA und BATANERO bereits in ihren einführenden Worten ihrer Studie *Judgments of Correlation in Scatter Plots* hin (vgl. ESTEPA & BATANERO 1996, p. 25 f). Die beiden spanischen Forscher untersuchten an 213 Oberstufenschülern (18 Jahre alt, je die Hälfte männlich bzw. weiblich), inwieweit sie die Beziehung zwischen zwei Variablen in Streudiagrammen erfassen können. Vorgelegt wurden ihnen je eine Aufgabe mit jeweiligem Diagramm, in dem (1.) keine erkennbare Beziehung zwischen den Größen bestand, (2.) eine negative Korrelation vorlag sowie (3.) ein direkter, positiver Zusammenhang vorhanden war (vgl. Abb. 2-43, S. 78).

Da laut CURCIO und PICHARD das *Lesen* und Interpretieren eines Graphs, insbesondere das Erfassen der Achsenbeschriftung und der Skalierung, mit eine der Grundvoraussetzungen für das Verständnis von Zusammenhängen ist, ließen ESTEPA und BATANERO bei allen Testaufgaben – noch vor der Frage nach Zusammenhängen – einzelne Werte aus den Plots rauslesen (MORITZ verfuhr in seiner Studie bei seiner Aufgabe mit der Anzahl an Schüler/inne/n pro Klasse und dem entsprechenden Lautstärkenlevel übrigens ebenso, vgl. Abb. 2-39 auf S. 72 „*Q2b*"). Die Ergebnisse decken sich mit den Erkenntnissen aus MORITZ' Studie: Die Schüler/innen hatten hier in der Regel keine Probleme, über 85 % (vgl. Abb. 2-44) bewältigten diesen Teil der Aufgabenstellung jeweils erfolgreich. Das Erfassen der Zusammenhänge bereitete jedoch vor allem bei Testaufgabe 3 einigen Probanden Schwierigkeiten: Nur knapp 22 % erfassten den direkten positiven Zusammenhang zwischen den beiden Größen (Studenten, die bei einem Prüfer besser bewertet wurden, erhielten tendenziell auch vom (unabhängigen) zweiten Prüfer ein besseres Urteil). ESTEPA und BATANERO ging es in ihrer Studie vor allem um die Vorgehensweise bzw.

Table 4. Frequency and percent of type of correlation assessed by the students

Item	Independence	Direct	Inverse	No answer
1	*177 (83.1)	8 (3.8)	24 (11.3)	4 (1.9)
2	19 (8.9)	11 (5.2)	*182 (85.4)	1 (0.5)
3	137 (64.3)	*46 (21.6)	18 (8.5)	12 (5.6)

*correct answer

Abb. 2-44: Einschätzungen der Oberstufenschüler bzgl. der im Diagramm dargestellten Korrelationen

um die Strategien, *wie* die Schüler/innen zu ihrer Einschätzung kamen, und ordneten die Antworten einem zehnstufigen Raster zu. Auffällig dabei war, dass sich selbst bei korrekt gelösten Aufgaben oft ein falsches Begründungsmuster dahinter verbarg (vgl. Abb. 2-45, S. 80).

Frequency and percent of strategies by implicit mathematical concept

Strategies	Item 1	Item 2	Item 3
GLOBAL	16(7.5)	3(1.4)	28(13.1)
INCREASE	15(7.0)	134(62.9)	8(3.8)
PATTERN	0	0	24(11.3)
CORRECTP	49(23.0)	27(12.7)	4(1.9)
INCORRECTP	4(1.9)	3(1.4)	25(11.7)
PTHEORY	22(10.3)	15(7.0)	6(2.8)
OTHERV	68(31.9)	1(0.5)	0
UNIFORM	14(6.6)	1(0.5)	21(9.9)
CAUSALITY	0	0	50(23.5)
OTHER	25(11.7)	29(13.6)	47(22.1)

Abb. 2-45: Die hier rot umrahmten Strategien wurden als „*incorrect*" klassifiziert.

Als Hauptgründe für die Anwendung einer falschen Strategie (trotz korrekter Antwort) bzw. die Falschbeantwortung einer Aufgabe machten ESTEPA und BATANERO zusammenfassend vor allem die folgenden drei Gründe verantwortlich:

1. Schüler/innen erwarten teilweise in der graphischen Darstellung den „perfekten" Zusammenhang, der die Beziehung zwischen den Variablen eindeutig und vollkommen darstellt. Ausreißer im Diagramm tolerieren sie nicht, ansonsten wird die Abhängigkeit der Größen in Frage gestellt.

2. Häufig ziehen die Schüler/innen nur einzelne Punkte eines Streudiagramms für ihre Begründung heran. Wenn diese in Werte in die vorgeschlagene Zusammenhangshypothese passt, reicht ihnen das als Begründung aus.

3. Schüler/innen greifen auf ihre grundsätzlichen Überzeugungen bzw. Erwartungen zurück, um einen Zusammenhang zwischen den dargestellten Größen zu begründen. Widerspricht die dargestellte Beziehung der Variablen dieser Erwartung, geben sie lieber an, es existiere überhaupt kein Zusammenhang zwischen den Größen.

→ *Vor allen der Punkt 1 und 2 sind im Hinblick auf die Forschungsfragen besonders interessant: Bestätigt sich in der vorliegenden Studie das* lokale Begründungsmuster? *Erkennen Schüler/innen den Zusammenhang zwischen zwei*

Größen, auch wenn dieser nicht beinahe perfekt *dargestellt ist? Aber auch Punkt 3 kann ggf. bei der Beantwortung der Testfragen eine entscheidende Rolle spielen: Widerspricht der dargestellte Sachverhalt der Erwartung des Probanden, könnte dieser u. U. seine Zusammenhangshypothese verwerfen.*

Dass sich Erwartung und Erfahrung auf das Erkennen eines Zusammenhangs in Daten auswirkt, konnten BATANERO und ESTEPA in einer weiteren Studie mit GODINO und GREEN verifizieren (N = 213), auch wenn in dieser Studie der Focus auf das Erkennen von Korrelationen in Kontigenztabellen gelegt wurde (BATANERO, ESTEPA, GODINO & GREEN 1996).

Obwohl die bisher angeführten Studien durchaus bemerkenswerte und für die vorliegende Forschungsarbeit relevante Aspekte liefern, darf nicht übersehen werden, dass sich alle geschilderten Ergebnisse vor allem auf die Darstellung und Auswertung grafischer Veranschaulichungen bivariater Datensätze beziehen. MORITZ bestätigt: *„Most research into the developing understanding of covariation has come from tasks involving graphs“* (MORITZ 2004, p. 232). Konkrete Forschungsergebnisse, inwieweit Schüler/innen Zusammenhänge zwischen zwei Größen aus Tabellen „lesen“ bzw. erfassen können, sind in der Literatur kaum zu finden. Es lassen sich zwar zumindest einzelne didaktische Kommentare und Handreichungen zum Thema Umgang mit Tabellen (z. B. bei der QUA-LiS NRW: „Strategien zum Umgang mit Tabellen“, 2014) finden – vor allem seit der nach PISA geäußerten Kritik am mangelnden Verständnis deutscher Schüler/innen diskontinuierlicher Texte – doch auch diese Abhandlungen gehen äußert rasch wieder in die Diskussion um das Verstehen grafischer Darstellungen über.

Da die hier vorgestellten Studien größtenteils zu ähnlichen Ergebnissen kommen, sind die vor allem für die vorliegende Forschungsarbeit relevanten Aspekte für einen besseren Überblick nochmals stichpunktartig zusammengefasst:

- Schüler/innen können in der Regel ohne Probleme Diagrammen konkret nachgefragte Werte entnehmen (MORITZ 2004, PEREIRA-MENDOZA 1995, PEREIRA-MENDOZA & MELLOR 1990, ESTEPA & BATANERO 1996).
- Scheinkorrelationen werden bereitwillig akzeptiert und selten kritisch hinterfragt (WATSON & MORITZ 1997).
- Positive Korrelationen werden tendenziell eher erkannt als negative (MORITZ 2004), allerdings erwarten die meisten Schüler/innen den „idealen" Zusammenhang; Ausreißer und eine zu große Streuung der Werte bringen unter Umständen die vorgebrachte Zusammenhangshypothese ganz zu Fall (ESTEPA & BATANERO 1996). Ebenso verhält es sich, wenn die entdeckte Korrelation den Erwartungen und Überzeugungen der Schüler/innen zuwiderläuft (ZIEFFLER 2008, ESTEPA & BATANERO 1996).
- Schüler/innen haben bei einem entdeckten Trend erhebliche Schwierigkeiten beim Einordnen neuer Werte in ein bestehendes Datengefüge; darüber hinaus erweisen sich die Begründungsmuster hierzu als besonders defizitär (MORITZ 2004, PEREIRA-MENDOZA 1995 und PEREIRA-MENDOZA & MELLOR 1990).
- Das Erkennen und Hinterfragen von Zusammenhängen bzw. Trends scheint abhängig vom Alter bzw. von der Jahrgangsstufe zu sein: Je älter die Probanden, desto erfolgreicher sind sie darin (MORITZ 2004, PEREIRA-MENDOZA & MELLOR 1990).

3 Einschränkungen und Fragestellungen

Allein aus den bisherigen Forschungsergebnissen ergäben sich unzählige weitere offene und vertiefende Fragestellungen, die es wert wären, ihnen in weiteren empirischen Untersuchungen nachzugehen. Schon die bloße Verifikation der Befunde des dargestellten Forschungsstands für den deutschsprachigen Raum wäre sicher ein lohnenswerter Forschungsschwerpunkt. Im Hinblick auf den Umfang dieser Arbeit müssen jedoch Einschränkungen hinsichtlich der Anzahl der zu untersuchenden Fragestellungen als auch Differenzierungen diesbezüglich vorgenommen werden, zumal es im natürlichen Interesse jedes Wissenschaftlers liegt, auch neuen, noch nicht vorgebrachten Hypothesen bzw. Theorien nachzugehen.

Die bereits im Kapitel 2.3 dargestellte mögliche Herangehensweise an einen bivariaten Datensatz zeigt, welch vielschichtiges Anforderungsniveau im Umgang mit bivariaten Datensätzen – und zwar bereits ab der Phase der tabellarischen Darstellung – erforderlich ist, um „im Interesse des betrachteten Sachproblems zu vertieften Erkenntnissen zu gelangen." (BIEHLER 1997, S. 8). Trotz der für sich einzeln betrachtet einfachen Techniken und mathematischen Elemente, ist die Bearbeitung bivariater Daten sowie die ständige Reflexion über die vorläufigen Ergebnisse während des gesamten Analyseprozesses ein insgesamt komplexer Vorgang, der sich wohl kaum in einer einzigen Studie untersuchen ließe. Viele Teildisziplinen, die einem die Explorative Datenanalyse abverlangt, wie z. B. das Sortieren und Aufbereiten der Daten, der kritische Einbezug des Sachkontexts oder gar eine rechnerische Approximation, können und sollen hier nicht geleistet werden: Nicht nur, weil sie das Anforderungsniveau und die (zeitliche) Leistungsgrenze der Schüler/innen in einem Setting übersteigen würden, sondern auch, weil aus wissenschaftlicher Sicht die Forderung nach validen und reliablen Ergebnisse eine Konzentration auf eine begrenzte Anzahl von Forschungsfragen nötig erscheinen lässt.

Wie im Kapitel 2.4 dargestellt, konzentriert sich die folgende Studie auf die Untersuchung des Umgangs mit bivariaten Daten, der sich auf das Auslesen und Interpretieren graphischer und tabellarischer Darstellungen inklusive einer Prognose bezieht. Der Schwerpunkt der Untersuchung fällt – wie ein Blick auf Forschungsfragen bereits verrät – eindeutig in den Bereich „*reading beyond the data*"

(F2 a - e), das *Lesen* über den bestehenden Datensatz hinaus. Vor allem die beiden Forschungsfragen 2 a und 2 b sollen dabei helfen, diese dritte Stufe einer Kompetenz im Umgang mit bivariaten Daten weiter auszudifferenzieren: Wie *gut* ordnen Schüler/innen neue Werte in das bestehende Datengefüge ein? Und wie tragfähig ist dann die entsprechende dazugehörige Begründung? Und: Haben Alter, Art der Korrelation (positiv/negativ) und die Darstellungsform (Graphik/Tabelle) Einfluss auf das erfolgreiche Bearbeiten von Aufgaben dieser Art?
Konkret soll nun im empirischen Teil der vorliegenden Arbeit folgenden Fragestellungen nachgegangen werden:

F1 a) Wie erfolgreich entnehmen Schüler/innen einer (bivariaten) tabellarischen Darstellung konkret nachgefragte Werte?

Reading the data

Die vorliegenden Forschungsergebnisse, insbesondere die Studien von Pereira-Mendoza und Mellor (1990) und Estepa und Batanero (1996), sprechen zumindest in Bezug auf Diagramme eine deutliche Sprache: Schüler/innen haben keine Schwierigkeiten, Diagrammen konkret nachgefragte Werte zu entnehmen (vgl. S. 79 - 81 und S. 83). Da in Tabellen die Daten in der Regel *direkt* vorliegen und nicht erst über Raster oder Hilfslinien herausgesucht und die Skalierungen der Achsen richtig interpretiert werden müssen, ist zu erwarten, dass die Erfolgsquote beim *Lesen* zusammengehörender Wertepaare in einer Tabelle mindestens genauso hoch sein wird bei Diagrammen.

F1 b) In welchem Ausmaß können Schüler/innen bei zweidimensionalen Tabellen Fragestellungen beantworten, die auf einen Vergleich zweier Merkmale zielen?

Reading between the data

Dieselben bereits vorliegenden Forschungsergebnisse von PEREIRA-MENDOZA und MELLOR (1990) hinsichtlich grafischer Darstellungen (vgl. S. 79 - 81) sollten sich auch in diesem Punkt eindeutig hinsichtlich tabellarischer Konstrukte bestätigen lassen: Fragen, die auf einen Vergleich innerhalb desselben Datensatzes abzielen (z. B. „Ein Storch hat durchschnittlich eine höhere Lebenserwartung als ein Huhn.“), sollten in hohem Maße korrekt beantwortet werden.

F2 a) Wie schlüssig ordnen Schüler/innen neue Werte in ein bestehendes Datengefüge ein? Welche Qualität hat dabei die dafür angegebene Erklärung?

Reading beyond the data

MORITZ (2004) konnte bereits zeigen, dass zumindest knapp 60 % der Probanden und vor allem die älteren Schüler seiner Stichprobe (7.- und 9.-Klässler) in der Lage sind, neue Werte in ein bereits bestehendes Datengefüge sinnvoll einzuordnen (S. 77 f: *Numerical Graph Interpretation*). PEREIRA-MENDOZA und MELLOR (1990) konstatierten eine deutlich pessimistischere Sichtweise (*„... the level of performance was extremely low“*, siehe S. 76), allerdings bestand die Stichprobe aus 4.- und 6.-Klässlern (prinzipiell belegen alle angeführten Studien, dass ältere Schüler/innen deutlich datenkompetenter auftreten). SHAUGHNESSY (2007) äußerte sich in seiner Zusammenfassung ähnlich, und auch ESTEPA und BATANERO (1996) zeigten sich beinahe überrascht, wie selten eine korrekte Einordnung mit entsprechender Begründung zusätzlicher Werte in ein bestehendes Datenkonstrukt gelingt. Das Erklärmuster, warum Schüler/innen einen Wert an einer bestimmten Stelle einordnen, ist in den entsprechenden Studien meist ähnlich und besteht größtenteils im Heranziehen der direkt benachbarten Punkte bzw. Werte. Besonders die Untersuchung von MORITZ (2004) belegt mit konkreten Beispielen, dass Schüler/innen oft nur *lokal* über den „Wert in der Mitte" argumentieren (vgl. S. 77 „*Numerical Graph Interpretation*").

Bei ESTEPA und BATANERO (1996) scheiterten vor allem jüngere Schüler an einer korrekten Einordnung neuer Werte, denn sie hielten Aufgaben dieser Art vielfach für nicht lösbar, da ja genau jener einzuordnende Wert in der Darstellung fehle.

Es ist folglich zu erwarten, dass die Teilnehmer entsprechend ihrer Altersstruktur (6. und 9. Jahrgangsstufe) in der vorliegenden Studie mehrheitlich neue Werte sinnvoll in einem Diagramm bzw. in einer Tabelle einordnen können, wobei sie vorrangig *lokal* argumentieren, d. h. mindestens einen oder zwei Nachbarwerte(paare) als Erklärung anführen.

F2 b) Inwieweit korreliert die Fähigkeit, neue Werte schlüssig in ein bestehendes Datengefüge einzuordnen, mit der Fähigkeit, einen Trend in Daten sachgerecht zu begründen?

Dieser Aspekt wurde in bisherigen Studien nicht untersucht. Man könnte mutmaßen, dass Schüler/innen, die neue Werte korrekt (d. h. innerhalb sinnvoller Grenzen) in ein bereits bestehendes Datengefüge einsortieren können, (gedanklich bereits vorher) auch den dahinterliegenden Trend identifiziert haben und dies entsprechend artikulieren können. Auch wenn das Erfassen des dahinterliegenden Trends laut MORITZ' (2004) Studie einigen Schüler/inne/n – es sind fast ausnahmslos die 9.-Klässler – gelingt (vgl. Abb. 2-41 auf S. 71), bedeutet das nicht automatisch, dass dies eine adäquate Begründung für die vorgebrachte Zusammenhangshypothese mit sich bringt. Ob nun zwischen einem hohen *Erklärlevel* und einem ausgeprägten *Begründungslevel* ein Zusammenhang besteht, soll diese Forschungsfrage klären.

F2 c) Inwiefern ist das Einordnen, Erklären und Begründen von Zusammenhängen abhängig vom Alter der Probanden?

Wie bereits bei F2 a) erwähnt, zeigt sich in beinahe allen Studien, dass unabhängig von der untersuchten Fähigkeit oder (Teil-) Kompetenz das Alter eine wesentliche Rolle zu spielen scheint. Hier ist zu erwarten, dass sich die Ergebnisse der bisherigen Untersuchungen bestätigen und die ältere Vergleichsgruppe der Studienteilnehmer (Schüler/innen der 9. Jgst.) deutlich besser abschneiden als ihre jüngeren Mitstreiter (6. Jgst.).

F2 d) Hängt das Erkennen eines Zusammenhangs von der Art der Korrelation ab? Sind Schüler/innen beim Einordnen neuer Werte bei einer zugrundeliegenden positiven Korrelation erfolgreicher als bei einer negativen Korrelation?

Der bisherige Forschungsstand lässt keine eindeutige Richtung für eine Beantwortung dieser Frage zu. Bei MORITZ (2004) lieferten die Probanden bei einer Aufgabe mit positiver Korrelation die besseren Ergebnisse (vgl. Abb. 2-37 (*negative association ↔ positive association*) auf S. 70), bei ESTEPA & BATANERO (1996) dagegen bei einer Aufgabe mit negativ korrelierten Datenpaaren (vgl. S. 78 inkl. Abb. 2-43 „*Independence, inverse, direct correlation*“). Anstelle der Wirkrichtung des Zusammenhangs scheint übereinstimmend vielmehr die Güte des Zusammenhangs (also ein hoher Korrelationskoeffizient) sowie die Übereinstimmung des Zusammenhangtyps mit der subjektiven Erwartung der Probanden zu sein.

Reading beyond the data

Da bei allen Aufgaben der vorliegenden Studie, denen eine Korrelation zugrunde liegt, der Betrag des Korrelationskoeffizienten bei mindestens 0,90 liegt, ist zu erwarten, dass die Schüler/innen die Aufgaben unabhängig von der Wirkrichtung des Zusammenhangs *gleich gut*, also ähnlich erfolgreich bearbeiten.

F2 e) Ist das erfolgreiche Einordnen neuer Werte in ein bestehendes Datengefüge abhängig von der Darstellungsform (Graph vs. Tabelle)?

Es gibt zwar Studien, deren Ergebnisse gewisse Aussagen über die Fähigkeiten beim Einordnen neuer Werte in einen bestehenden Datensatz ermöglichen (vgl. Ausführungen zur Forschungsfrage 2 a), jedoch beziehen sich alle geschilderten Ergebnisse auf die Darstellung und Auswertung *grafischer Veranschaulichungen* bivariater Datensätze. Untersuchungen, wie erfolgreich das Schüler/inne/n anhand von *Tabellen* gelingt, sind bis dato nicht bekannt. Auch ohne Grundlage empirischer Studien wird man intuitiv geneigt sein, die Vermutung zu äußern, dass eine Einordnung neuer Werte leichter gelingt, wenn sie graphisch aufbereitet zur Verfügung stehen (vgl. z. B. BIEHLER & SCHWEYNOCH 1999, S. 18). Dass man in Tabellen – vor allem in großen und noch dazu unsortierten Tabellen – beim ersten Betrachten ein gewisses Muster ins Auge fällt, scheint doch sehr unwahrscheinlich. Diese Vermutung lässt sich zumindest anhand der Mathematikdidaktik bzw. der Statistikratgeber untermauern: Übereinstimmend gilt die Empfehlung, vor allem bei größeren Datensätzen rasch zu einer graphischen Darstellung überzugehen, denn „sie verspricht mehr Übersicht und damit die größere Chance, Trends oder andere Besonderheiten zu entdecken." (NOLL & SCHMIDT 1994, S. 52). Selbst bei einer rein computergestützten, rechnerischen Aufbereitung von

Daten rät Cleff, noch *vor* der Berechnung gewisser Zusammenhangsmaße, Daten stets erst zu visualisieren, um einen ersten Eindruck über die Wirkungsrichtung, die Form und die Stärke eines möglichen Zusammenhangs zu gewinnen. Da die Tabellen in der vorliegenden Studie relativ überschaubar und zudem bereits nach einem Merkmal vorsortiert sind, ist das Erkennen eines Trends und die damit gefragte Einordnung neuer Werte sicherlich ohne größere Schwierigkeiten zu bewerkstelligen, jedoch ist aufgrund der „Kraft der Bilder“ (der Eingängigkeit von Diagrammen) zu erwarten, dass das Einordnen neuer Werte in ein bestehendes Datenkonstrukt sich erfolgreicher gestaltet, wenn die Daten graphisch aufbereitet als Diagramm vorliegen.

Reading beyond the data

4 Methodischer Rahmen

Dieses Kapitel beschreibt schwerpunktmäßig die Überlegungen und das Konzept des Testhefts samt seiner Aufgaben (Testinstrument) sowie die anschließenden Prinzipien zur Auswertung (Kodierung), die die Voraussetzung und Grundlage für die Untersuchung der Forschungsfragen bilden. Vorab wird auf den Erhebungsrahmen (Stichprobe) eingegangen.

4.1 Stichprobe

Die Studie wurde im Juli 2016 an drei bayerischen Realschulen unterschiedlicher regionaler Struktur (Landeshauptstadt, kleinere bayerische Großstadt und eine Kleinstadt der ländlichen Region) durchgeführt. Es nahmen insgesamt N = 160 Schüler/innen teil, davon waren 33 männlich und 137 weiblich. An jeder dieser drei Schulen wurde jeweils eine 6. und eine 9. Klasse ausgewählt. Das durchschnittliche Alter der 6.-Klässler ($N_{6.\text{-Kl.}}$ = 83) lag bei 12,0, das der 9.-Klässler ($N_{9.\text{-Kl.}}$ = 77) bei 15,1 Jahren (Abb. 4-1)

	Realschule			
	Landeshauptstadt (> 1 Mio Einwohner)	Großstadt (> 100 000 Einwohner)	Kleinstadt (< 20 000 Einwohner)	
6. Klasse	31	27	25	**83**
9. Klasse	22	25	30	**77**
N	**53**	**52**	**55**	**160**

Abb. 4-1: Teilnehmerzahlen (absolute Angaben) an den drei Realschulen

Die einführenden Hinweise und die Durchführung übernahm jeweils die vor Ort unterrichtende Mathematiklehrkraft. Die effektive Bearbeitungszeit war auf 35 Minuten angesetzt, so dass die Durchführung in einer regulären Unterrichtsstunde ohne größeren Organisationsaufwand möglich war.

4.2 Testinstrument

Diese Studie folgt vorrangig dem quantitativen Forschungsansatz. Sie hat – trotz einer überschaubaren Stichprobengröße – das Ziel, tendenziell verallgemeinerbare Aussagen hervorzubringen und vorab formulierten Fragestellungen nachzugehen. Beides sind zwar typische „Charakteristika der quantitativen Forschung" (BURZAN 2005, S. 25), allerdings verschließt sich die Entscheidung für den quantitativen Ansatz zur Hypothesenüberprüfung nicht automatisch einer Hypothesengenerierung (vgl. BURZAN 2005, S. 25 f): Das vorliegende Aufgabensetting enthält auch Freitextantworten („erkläre", „begründe") und weist Elemente qualitativer Sozialforschung auf, um eine gewisse Tiefe der Forschungsfragen erfassen zu können.

Die Konzeption des Fragebogens berücksichtigt im Aufbau bzw. in seiner grundsätzlichen Gestaltung – soweit möglich – fast alle Kriterien, die an einen „guten" Fragebogen zu stellen sind (vgl. STIER 1996, S. 180 f und S. 183 ff):

- einfache, altersadäquate Wortwahl; angepasstes Sprachniveau; Vermeidung von Fachausdrücken (soweit möglich)
- neutrale (nicht präjudizierende) und eindeutige Begriffe
- Vermeidung von Suggestivfragen und doppelter Verneinungen
- Ausgewogenheit der positiven und negativen Antwortmöglichkeiten („formal balanciert")
- klare Gliederung/Strukturierung
- angemessene Ausfülldauer, Zusicherung der Anonymität, Dank für die Mitarbeit

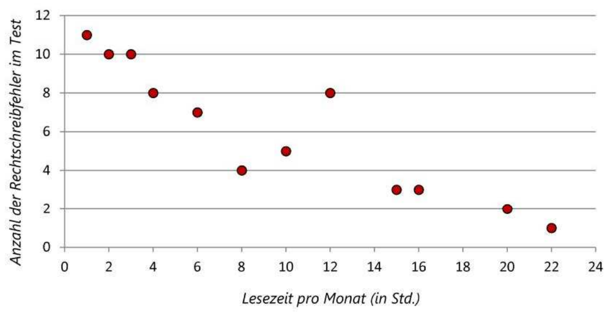
Für die Beantwortung der Fragestellungen wurde ein Testheft mit insgesamt zehn Aufgaben konzipiert. Der Übersichtlichkeit wegen beginnt – mit Ausnahme der beiden ersten Aufgaben – jede Seite mit einer neuen Aufgabe, so dass lästiges Hin- und Herblättern bei der Beantwortung der Fragen mithilfe des jeweiligen Diagramms bzw. der jeweiligen Tabelle entfällt. Jede Aufgabe – ebenfalls mit Ausnahme der ersten beiden Aufgaben – ist hinsichtlich ihrer Antwortmöglichkeiten (Anzahl und Systematik der Ankreuzmöglichkeiten, Reihenfolge der Fragestellung)

gleich gehalten, was eine raschere Bearbeitung des Testhefts vor allem nach den ersten Aufgaben ermöglichen sollte.
Um vergleichbare und verwertbare Aussagen zu erhalten, war es unumgänglich, bei den Fragen zur „Einordnungskompetenz" auf eine stets gleichlautende Formulierung zurückzugreifen („Welchen Wert vermutest du …?"). Auch die sich anschließende Erklärungsaufforderung („Erkläre möglichst genau, wie du auf diesen Wert kommst.") war in dieser stets gleichbleibenden Formulierung für eine hinreichende Auswertung der Forschungsfragen notwendig. Diese Monotonie ließe ein gewisses Reaktanzverhalten erwarten, welches sich bei der Auswertung allerdings nicht bestätigen ließ (vgl. BURZAN 2005, S. 74).

Die zehn Aufgaben schildern jeweils kurz einen Sachverhalt und liefern dazu einen Datensatz in tabellarischer oder grafischer Form (Streudiagramm). Bei acht dieser Aufgaben lässt sich zwischen den dargestellten Größen ein Zusammenhang nachweisen bzw. soll dieser bei näherer Betrachtung ersichtlich werden; bei zwei Aufgaben korrelieren die in der Aufgabenstellung benannten Merkmale nicht. Die Abbildung 4-2 zeigt exemplarisch den Auszug einer Aufgabe, bei der der Zusammen-

5. Lesen und Rechtschreibung

Zwölf Schüler wurden *vor* einem Rechtschreibtest befragt, wie viel Stunden sie ungefähr pro Monat lesen. Anschließend wurde der Test ausgewertet und das Ergebnis grafisch dargestellt.



☞ Schau dir das Diagramm in Ruhe an und überlege:

5 a) Ein anderer Schüler, der bei diesem Test nicht mitgeschrieben hat (er war an diesem Tag leider krank), gibt an, dass er ca. 11 Stunden pro Monat liest. Was glaubst du, wie viele Rechtschreibfehler er wohl haben würde, wenn er auch diesen Test schreibt?

Antwort:

Erkläre, wie du auf diese Anzahl kommst.

Abb. 4-2: Auszug aus Aufgabe 5 („Lesen und Rechtschreibung") des Testhefts hang über ein Streudiagramm erfasst werden kann, die Abbildung 4-3 den Auszug einer Aufgabe mit tabellarisch vorgegeben Daten.

6) Wer ist der Schnellste?

In einem Reaktionstest wird auf einem Bildschirm eine Farbe angezeigt. Die Testperson soll dann so schnell wie möglich den Knopf mit der richtigen Farbe drücken. Erst wenn der richtige Knopf gedrückt wurde, wird die nächste Farbe angezeigt.

Die nebenstehende Tabelle zeigt, wie lange Testpersonen verschiedenen Alters durchschnittlich gebraucht haben, um den richtigen Knopf zu finden.

	Reaktionszeit in Sekunden	**Alter** in Jahren
Testperson 1	0,40	12
Testperson 2	0,45	16
Testperson 3	0,50	16
Testperson 4	0,55	17
Testperson 5	0,60	20
Testperson 6	0,65	26
Testperson 7	0,65	38
Testperson 8	0,75	32
Testperson 9	0,80	40
Testperson 10	0,90	45
Testperson 11	1,00	35
Testperson 12	1,30	60
Testperson 13	1,30	72
Testperson 14	1,35	55

☞ Schau dir die Tabelle an und überlege:

6 a) Eine weitere Testperson im Alter von 42 Jahren macht denselben Reaktionstest. Was schätzt du, wie viele Sekunden diese Person durchschnittlich braucht um den richtigen Knopf zu drücken?

Antwort:

Erkläre, wie du auf diesen Wert kommst.

Abb. 4-3: Auszug aus Aufgabe 6 („Wer ist der Schnellste?“) des Testhefts

Bei der Konzeption des Testhefts wurde darauf geachtet, dass bei den Aufgaben positive und negative Korrelationen ebenso häufig auftreten wie auch die jeweilige Darstellungsform der Daten: je nach Korrelationstyp finden sich deshalb je zwei Aufgaben, bei denen die Daten als Diagramm dargeboten werden, und je zwei Aufgaben, bei denen das Datenmaterial in einer Tabelle präsentiert wird.

Die Abbildung 4-4 (S. 94) zeigt einen tabellarischen Überblick hinsichtlich des zugrundeliegenden Korrelationstyps, der dargebotenen Präsentationsform des Datenmaterials und des möglichen Antworttyps bei der Frage nach erkannten Zusammenhängen.

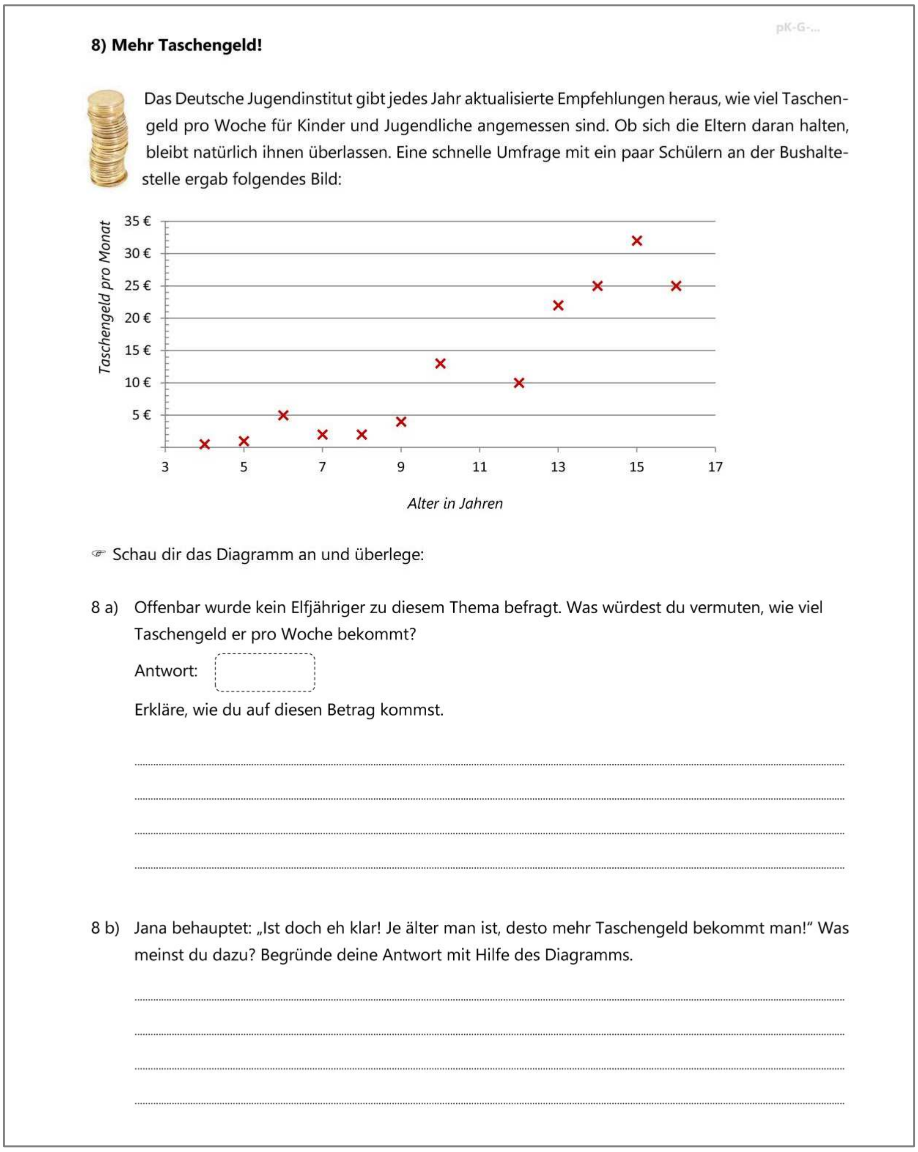
Korrelations-typ	Darstellungsform	Antworttyp	Bemerkung	Aufgaben-nr.
positiv	Tabelle	offen (Freitextantwort)		10
positiv	Tabelle	geschlossen (Ankreuzen)		6
positiv	Diagramm	offen (Freitextantwort)		8
positiv	Diagramm	geschlossen (Ankreuzen)		5
negativ	Tabelle	---	ausführlich	1
negativ	Tabelle	geschlossen (Ankreuzen)	ausführlich	2
negativ	Diagramm	offen (Freitextantwort)		4
negativ	Diagramm	geschlossen (Ankreuzen)		7
ohne	Diagramm	offen (Freitextantwort)		9
ohne	Diagramm	geschlossen (Ankreuzen)		3

Abb. 4-4: Überblick über die Aufgabentypen im Testheft

Alle Aufgaben, denen im Datenmaterial eine Korrelation zugrunde liegt (also bei acht der zehn Aufgaben), lassen sich prinzipiell in zwei Teile gliedern:

- Im ersten Teil – in der Regel die Teilaufgabe a) – ist der Tabelle oder dem Streudiagramm ein fiktives Wertepaar einzuordnen: Dabei ist der Wert des ersten Merkmals vorgegeben, der (korrelierende) Wert des zweiten Merkmals wird erfragt. Bei der Aufgabe 5 („Lesen und Rechtschreibung“, Abb. 4-2 auf S. 92) ist bei der Teilaufgabe a) z. B. die angenommene Lesezeit mit 11 Stunden pro Monat vorgegeben und der Proband soll die vermutete (modellhaft angenommene) Fehleranzahl im Rechtschreibtest angeben. Bei jeder dieser Aufgaben wird eine Erklärung für den angegeben Wert gefordert („Erkläre, wie du auf diesen Wert kommst.“).
  Dieser Aufgabentypus wird im weiteren Verlauf auch als „Erkläraufgabe“ und der erreichte Score als „Erklärlevel“ (kurz E-Level) bezeichnet (vgl. Kap. 4.3).

- Im zweiten Teil – in der Regel die Teilaufgabe b) – sollten die Schüler/innen zu einer Aussage bzw. Behauptung über den im Diagramm bzw. in der Tabelle gemachten Sachverhalt Stellung beziehen: Entweder durch Ankreuzen oder als frei zu formulierende Antwort. Da die Auswahlmöglichkeiten bei den An-kreuz-aufgaben die gemachte Aussage bzw. Behauptung begründen oder direkt nach der Begründung als frei zu formulierende Antwort gefordert wurde („Begründe deine Antwort mit Hilfe des Diagramms.“, vgl. Abb. 4-5 auf S. 95), werden diese Aufgaben im weiteren Verlauf auch „Bergündungsaufgaben“ und der erreichte Score als „Begründungslevel“ (kurz B-Level) bezeichnet.



8) Mehr Taschengeld!

Das Deutsche Jugendinstitut gibt jedes Jahr aktualisierte Empfehlungen heraus, wie viel Taschengeld pro Woche für Kinder und Jugendliche angemessen sind. Ob sich die Eltern daran halten, bleibt natürlich ihnen überlassen. Eine schnelle Umfrage mit ein paar Schülern an der Bushaltestelle ergab folgendes Bild:

☞ Schau dir das Diagramm an und überlege:

8 a) Offenbar wurde kein Elfjähriger zu diesem Thema befragt. Was würdest du vermuten, wie viel Taschengeld er pro Woche bekommt?

Antwort:

Erkläre, wie du auf diesen Betrag kommst.

8 b) Jana behauptet: „Ist doch eh klar! Je älter man ist, desto mehr Taschengeld bekommt man!“ Was meinst du dazu? Begründe deine Antwort mit Hilfe des Diagramms.

Abb. 4-5: Aufgabe „Mehr Taschengeld!“ aus dem Testheft als Beispiel für eine „Begründungsaufgabe“ (8 b)

Die ersten beiden Aufgaben des Testhefts sind absichtlich mit je einer Doppelseite (vgl. Abb. 4-6, S. 96) ausführlicher gestaltet; sie dienen im didaktischen Sinne der Heranführung an die Thematik und die Vorgehensweise für die weiteren Aufgaben im Testheft. Dass Schüler/innen Diagrammen und Tabellen korrekt und zuverlässig Werte entnehmen und vergleichen können, wurde aufgrund des bekannten

A1) Lebenserwartung und Herzfrequenz von Tieren

Wenn in der folgenden Aufgabe z. B. eine *durchschnittliche* Herzfrequenz von 40 angegeben ist, dann bedeutet das, dass das Herz dieses Tieres normalerweise 40× pro Minute schlägt.
(Natürlich kann das Herz auch mal etwas schneller oder langsamer schlagen, denn auch ein Tier kann mal aufgeregt sein oder muss ein wenig schlafen.)

Ebenso verhält es sich mit der *durchschnittlichen* Lebenserwartung: Ein Löwe z. B. hat nicht immer eine exakte Lebenserwartung von 23 Jahren. Manche Löwen werden ein bisschen älter, manche versterben ein wenig früher. Aber im Schnitt werden sie 23 Jahre alt.

	Tier	durchschnittliche Anzahl der Herzschläge (pro Minute)	durchschnittliche Lebenserwartung (in Jahren)
1	Löwe	40	23
2	See-Elefant (an der Wasseroberfläche)	60	20
3	Ziege	68	18
4	Schaf	70	16
5	Storch	85	18
6	Fuchs	100	13
7	Hauskatze	120	12
8	Kaninchen	215	9
9	Meerschweinchen	256	10
10	Huhn	353	6
11	Krähe	380	3
12	Goldhamster	425	2

☞ Schau dir nun in Ruhe die obige Tabelle an beantworte mit Hilfe dieser Tabelle folgende Fragen:

1 a) Das Herz eines Kaninchens schlägt durchschnittlich 215 Mal pro Minute.
☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

1 b) Schreibe in das Kästchen, welche Lebenserwartung ein Fuchs durchschnittlich hat:

Seite 2

1 c) Ein Storch hat eine durchschnittlich höhere Lebenserwartung als ein Huhn.
☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

1 d) Die durchschnittliche Herzfrequenz eines Meerschweinchens ist höher als die eines Goldhamsters.
☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

1 e) Eine Giraffe hat eine durchschnittliche Herzfrequenz von 75.
Was vermutest du, wie alt so eine Giraffe wird?

Antwort:

Erkläre möglichst genau, wie du auf diese Zahl kommst.

→ *Bitte umblättern!*

Seite 3

Abb. 4-6: Die ausführliche Aufgabe 1 erstreckt sich im Testheft über eine Doppelseite

8	Erziehung und Unterricht	3000	4
9	freiberufl., techn. oder wiss. Dienstleistungen	3500	4
10	Energieversorgung	4000	1
11	Information und Kommunikation	4700	2
12	Finanzen und Versicherung	5700	1

☞ Schau dir erst wieder die obige Tabelle an beantworte dann mit Hilfe der Tabelle folgende Fragen:

2 a) Personen, die im Gesundheitswesen angestellt sind, verdienen durchschnittlich 2700 € pro Monat.
☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

2 b) Schreibe in das Kästchen, wie viele Krankheitstage eine Person aus dem Bereich „Information und Kommunikation" durchschnittlich pro Jahr hat:

2 c) Personen aus dem Bereich „Erziehung und Unterricht" verdienen durchschnittlich weniger als Personen, die im Gesundheitswesen beschäftigt sind.
☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

2 d) Personen aus dem Baugewerbe sind durchschnittlich öfter krank gemeldet als Personen aus dem Gastgewerbe?
☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

Abb. 4-7: Auszug aus der ausführlichen Aufgabe 2 mit Elementen zu *reading the data* und *reading between the data* (Teilaufgaben a) bis d))

Forschungsstands (vgl. Kap. 2.5) prinzipiell nicht in Frage gestellt. Die diesbezüglichen Aufgaben 1 a) - d) sowie 2 a) - d) (vgl. Abb. 4-7) sollen u. a. die bereits bekannten Befunde für den deutschsprachigen Raum bestätigen. In Hinblick auf das Niveau des Fragebogens und den zur Verfügung stehenden zeitlichen Bearbeitungsrahmen sind deshalb ab Aufgabe 3 auch keine weiteren Items zu *reading the data* und *reading between the data* mehr enthalten. Darüber hinaus darf das erfolgreiche Auswerten und Beantworten der jeweils ersten vier Teilaufgaben bei Aufgabe 1 und 2 auch als motivierende und Selbstvertrauen stärkende Einführung angesehen werden, wie sie auch Stier in seinen Empfehlungen zum Fragebogendesign unter „Einleitungs- und Übergangsfragen" (Stier 1996, S. 183) beschreibt. Aus demselben Grund fehlt bei der allerersten Aufgabe der offene Antworttyp bei Aufgabe 1 (das Betrachten der obigen Tabelle (Abb. 4-4, S. 94) lässt zu Recht vermuten, dass je fünf Mal der offene und je fünf Mal der geschlossene Antworttypus vorkommen sollte). Ein zu frühes Einsteigen mit aufwändigen und für die Schüler/innen kompliziert anmutenden Begründungen sollte durch das vermeintliche Fehlen einer Aufgabe 1 f vermieden werden.

Die Aufgabe 2 f) soll in diesem Sinne mit einer Möglichkeit, korrekte Aussagen anzukreuzen, die angeratene Überleitung darstellen, um mit dem Modus für die kommenden Aufgaben vertraut gemacht zu werden (Abb. 4-8).

2 f) Kreuze an, welche Aussagen du mit Hilfe des Diagramms machen kannst:

☐ Je mehr man durchschnittlich verdient, desto häufiger ist man krank.
☐ Wer mehr verdient, ist seltener krank.
☐ Je mehr Krankentage man durchschnittlich hat, desto weniger verdient man.
☐ Zwischen Einkommen und Krankentagen besteht kein Zusammenhang.

Abb. 4-8: Die erste „Begründungsaufgabe" des Testhefts

Für eine bessere Vergleichbarkeit hinsichtlich des Erfassens des abgefragten Korrelationstyps und des Einordnens neuer Werte in das bestehende Datengefüge wurden in der Regel bei allen Aufgaben zwischen 11 und 15 Wertepaare vorgegeben. Jeder Datensatz enthält mindestens einen deutlichen Ausreißer (vgl. Abb. 4-9, S. 98): Das ermöglicht zum einen zu überprüfen, ob dadurch das Vertrauen in die Zusammenhangshypothese geschmälert wird (vgl. die Ergebnisse von

ESTEPA und BATANERO (S. 80), die feststellten, dass Schüler/innen meist den idealen Zusammenhang (einen Korrelationskoeffizienten nahe 1) erwarten), zum anderen lässt sich dadurch eindeutiger feststellen, auf welche Weise Schüler/innen bei der Einordnung neuer Werte in das bestehende Datengefüge vorgehen: Orientieren sie sich lediglich an den direkt benachbarten Wertepaaren (lokale Sicht), oder haben sie den Gesamttrend im Auge (globale Sicht) und können sich vom Ausreißer gedanklich lösen bzw. die Trendlinie glätten?

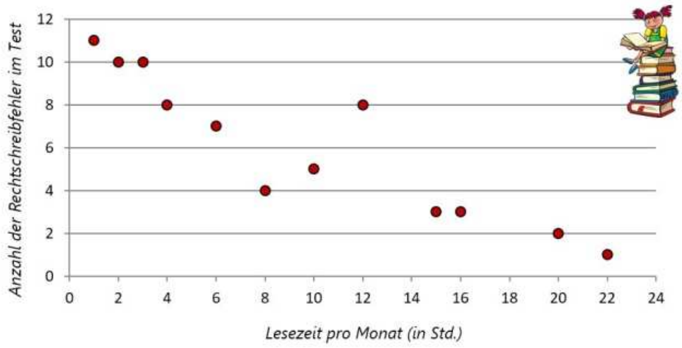


Abb. 4-9: Beispiel für einen typischen Ausreißer: Hier das Wertepaar (12|8) bei Aufgabe 5 des Testhefts

Um das Erkennen des zugrunde gelegten Korrelationstyps in den Tabellen zu erleichtern, wurden die Wertepaare bereits nach einem Merkmal sortiert abgedruckt (vgl. Kap. 3 „Einschränkungen ...“). Vermutlich gelänge das Erfassen eines Trends in tabellarisch dargestellten Daten bei einer solch überschaubaren Datensatzgröße auch bei einer nicht sortierten Vorgabe, allerdings wäre diese Leistung dann von Studienteilnehmern selbst zu erbringen, was einerseits nicht zwangsläufig jedem Probanden gelingen und andererseits einen deutlichen höheren Zeitaufwand bedeuten würde.

Ohnehin ist das Testheft insgesamt hinsichtlich der Anzahl der Aufgaben als auch der Fülle der zu verarbeitenden Texte für einen 35-minütigen Test am oberen Konzentrations- und Leistungslimit der Probanden anzusiedeln. KRAUTH empfiehlt selbst für „gesunde Erwachsene“, dass eine Bearbeitungszeit von 30 Minuten nicht überschritten werden sollte; bei Kindern sollte sie sogar „erheblich unter dieser Zeit“ liegen (KRAUTH 1995, S. 44).

4.3 Kodierung

Bei acht der zehn Aufgaben des Testhefts war es die Aufgabe der Probanden, einen weiteren vorgegebenen, noch nicht im Diagramm oder der Tabelle dargestellten Wert, einzuordnen (*reading beyond the data*). Die korrekte Angabe eines Werts innerhalb eines vorab festgelegten Intervalls (siehe Anhang 9.2 ab S. 146) wurde – eine anschließende entsprechende Erklärung (s. u.) vorausgesetzt – mit einer gewissen Punktzahl *vergütet*; die bloße Angabe eines Zahlenwerts – auch wenn er innerhalb des gültigen Intervalls lag – wurde mit 0 Punkten gewertet. Das Gültigkeitsintervall wurde in der Regel über die Ermittlung des *optimalen* Wertes (rechnerisch ermittelt über eine Regressionsfunktion) und (subjektiv) sinnvoller Ober- und Untergrenzen (meist die direkt benachbarten Wertepaare) festgesetzt. Da die anschließenden Erklärungen, wie die Schüler/innen auf diesen Wert gekommen sind, frei formuliert werden konnten, wurden Kategorien festgelegt, um die Antworten klassifizieren zu können. Dazu wurden die Antworten in fünf Abstufungen entsprechenden *Erklärleveln* (E0 bis E4) zugeordnet, die sich wie folgt darstellen:

E0: Es wurde keine Angabe gemacht, der angegebene Wert liegt nicht innerhalb der tolerierten Intervallgrenzen oder die Erklärung für den gefundenen Wert ist falsch, nicht sinnvoll oder nicht nachvollziehbar.

Erklärungen, die den entsprechenden Datensatz ignorieren und/oder auf einer bloß subjektiven Sicht gründen, wurden hier eingeordnet. Beispielantworten*:

„Habe ich im Tierpark auf einem Schild gelesen.“
„Ich denke es könnte stimmen, weil es mit Tieren so sein kann.“
„Man kann oft krank sein, egal wie viel man verdient.“

Angaben wie *„Das kann man nicht sagen.“* wurden ebenfalls der Stufe E0 zugeordnet; zumal bei diesem Antworttypus dann ohnehin kein Wert angegeben wurde.

* *Alle Schülerantworten wurden unverändert inkl. etwaiger Rechtschreib- und Satzzeichenfehler übernommen.*

E1: Der angegebene Wert ist korrekt bzw. nachvollziehbar, aber die angegebene Erklärung fehlt, ist falsch oder bezieht sich nicht die vorliegenden Daten. Beispielantworten für dieses Level:

„Ich habe geraten."
„Ich habe mich an der Tabelle orientiert."

E2: Es wurde ein korrekter/sinnvoller Wert angegeben und die dazugehörige Erklärung orientiert sich am gegebenen Datensatz. Allerdings orientiert sich die Einordnung offensichtlich nur an einem benachbarten Datenpaar. Beispielantworten:

lokal einseitig

„Die Klasse 6 E hat 23 Schüler und die sind 50 dB laut."
„Die Testperson 9 hat eine Reaktionszeit von 0,80 s."

E3: Der angegebene Wert ist korrekt bzw. sinnvoll ins bestehende Datengefüge eingeordnet. Die entsprechende Erklärung erfolgt anhand der vorliegenden Daten, wobei die Einordnung über die beiden benachbarten Wertepaare erfolgt. Beispielantworten auf diesem Level:

lokal zweiseitig

„Liegt zwischen 22 und 24 Schülern."
„Eine Person mit 40 braucht 0,8 s, eine mit 45 0,9 s. Der Wert sollte dazwischen liegen."
„Ich habe zwischen 2001 und 2003 verglichen."

E4: Der angegebene Wert ist korrekt bzw. sinnvoll ins bestehende Datengefüge eingeordnet. Die entsprechende Erklärung weist auf das Erfassen der Gesamttendenz hin.

global

„Je weniger das Herz schlägt, desto älter wird das Lebewesen."
„Wer mehr liest, macht weniger Fehler."

Dieses fünfstufige, hierarchische Modell ermöglicht in dieser Form eine differenzierte Auswertung sowie eine praktikable Bewertungsmöglichkeit: Pro Erklärlevel wird ein entsprechender Score vergeben, z. B. Erklärlevel E2 $\Rightarrow$ 2 Punkte, Erklärlevel E3 $\Rightarrow$ 3 Punkte, etc.

Der obigen Klassifizierung der Antwortkategorien kann man entnehmen, dass die Level E2 bis E4 die Angabe eines korrekten Wertes inklusive einer nachvollziehbaren Erklärung voraussetzen. Allerdings dürfen diese Abstufungen nicht automatisch als eine Art qualitative Abstufung von Datenkompetenz (hier die Stufe *reading beyond the data*, vgl. Kap. 2.4: „Die Sicht dieser Studie auf Datenkompetenz …“) aufgefasst werden: Die bloße korrekte Einordnung inklusive einer lokalen Argumentationsweise muss nicht bedeuten, dass ein Proband den Trend in den Daten erfasst hat (vgl. Forschungsfrage 2 a, ab S. 111). Diese Abstufungen werden diesbezüglich erst im Zusammenspiel mit den Ankreuzaufgaben (z. B. „Kreuze an, welche Aussagen du mit Hilfe der Tabelle machen kannst.“) und den *Begründungslevels* (siehe unten) aussagekräftig.

Bei vier der zehn Aufgaben sollten die Schüler/innen mit eigenen Worten zudem begründen, ob sie einen Zusammenhang zwischen den dargebotenen Merkmalen erkennen. Bei zwei dieser Aufgaben war zu einer bereits vorgegebenen Behauptung Stellung zu nehmen („Jana behauptet: » ... Je älter man ist, desto mehr Taschengeld bekommt man!« “, vgl. Abb. 4-5 auf S. 95), bei den anderen beiden konnten sich die Testteilnehmer frei äußern („Erkennst du ... einen Zusammenhang ...? Begründe deine Antwort“, vgl. Aufgabe 10 b) im Testheft (Anhang S. 145)).

Auch hier mussten die Schülerantworten erst kategorisiert werden: Ähnlich den Stufen der *Erkläraufgaben* wurden bei den Antworten zu diesen *Begründungsaufgaben* ebenfalls fünf Abstufungen (= „Begründungslevel“) vorgenommen. Die gewählte Parallelität bzgl. der Anzahl und Hierarchie der Erklärlevel erfolgte dabei ganz bewusst, um für die spätere Auswertung eine zweckdienliche Vergleichsbasis zu schaffen. Ähnlich wie bei den Erklärleveln spiegeln die Begründungslevel von B0 (Nichtbearbeitung der Teilaufgabe bzw. Fehlen einer qualifizierten Begründung) bis B4 (sachgerechte Begründung anhand des vorliegenden Datenmaterials) die Bandbreite aller schriftlich fixierten Schülerantworten wider.

Konkret erfolgte die Klassifizierung nach den folgenden Kriterien:

B0: Die Teilaufgabe wurde nicht bearbeitet bzw. keine Antwort gegeben. //
Eine Korrelation wurde nicht oder falsch erfasst (bzw. das Fehlen der Korrelation bei Aufgabe 9). //
Beispielantworten[*] für die Einordnung in dieses Level:

„Manche Eltern erlauben mehr und andere weniger Taschengeld.“
„Die Gene sind verantwortlich.“
„Nein, das kann man nicht berechnen.“
„Nein weil das unlogisch ist.“

B1: Die Korrelation wird erkannt und die Frage (korrekt) beantwortet, allerdings ohne weitere Begründung bzw. fußend auf einer rein subjektiven Sichtweise. Es erfolgt kein Bezug zu den Daten. Beispielantworten für dieses Level:

„Ja, das stimmt.“
„Ja weil ältere Kinder mit Geld etwas anfangen können.“
„Ja stimmt, je älter man wird desto öfter und mehr braucht man.“

Die Zustimmung bzw. Ablehnung zu einer vorgegebenen Aussage musste ersichtlich sein, sonst wurde die Antwort bei B0 eingestuft.

B2: Die Korrelation wird erkannt bzw. der vorgegebenen Behauptung zugestimmt, auf dieser Stufe allerdings mit eindeutigem Hinweis auf das Diagramm bzw. die Tabelle. Beispielantworten:

pauschaler Datenbezug

„Ja, anhand des Diagramms kann man das erkennen.“
„Das Diagramm sagt das zwar und somit gebe ich Max recht ...“
„Max hat recht, denn das liest man von dem Diagramm heraus, aber für mich selber ist das unlogisch ...“

---

[*] *Alle Schülerantworten wurden unverändert inkl. etwaiger Rechtschreib- und Satzzeichenfehler übernommen.*

B3: Die Korrelation wird erfasst, wobei zusätzlich eine konkrete Auseinandersetzung mit den Daten erfolgt. Die Zustimmung oder Abweichung vom angedachten Trend erfolgt mit einer sachgerechten Begründung anhand des Diagramms bzw. der Tabelle mithilfe mindestens eines Wertepaares.

konkreter Datenbezug

Beispielantworten:

„Nicht unbedingt, denn z. B. die Klasse 6 B und 6 E sind gleich laut."

„Das glaube ich nicht, weil auch mit wenigen Leuten es sehr leise sein kann. Zum Beispiel zw. 22 und 24 Schülern da ist der Lärmpegel nur auf 50."

Das teilweise bzw. „streckenweise" Erfassen einer Tendenz anhand einiger weniger Wertepaare sowie das Erkennen einer Tendenz an bloß einem Merkmal wurde ebenfalls hier eingeordnet. Beispiele:

„Ja, zwischen 1 °C und 12 °C trifft das zu."
„Die Lautstärke wird immer leiser."

B4: Korrektes Erfassen der Korrelation und entsprechende Beschreibung. Die Begründung bezieht die Sicht auf das Ganze (inkl. Nennung der entsprechenden Merkmale) oder zumindest mehrere Wertepaare mit ein. Beispielantworten:

umfassender Datenbezug

„Desto mehr Konsum von Tiefkühlkost, desto mehr Strom."

„Ich stimme Max zu, weil man ja in dem Diagramm sieht, dass bei 24 Schülern der Lärmpegel am höchsten ist und dann bei 26, 28 und 30 Schülern wird der Lärmpegel immer tiefer."

„In dem Diagramm bekommen zwar übermäßig ältere mehr Taschengeld, jedoch kann man das nicht allgemein sagen, da z. B. im Diagramm ein 14-jähriger genauso viel bekommt wie ein 16-jähriger."

Für die weitere Auswertung und Beantwortung der Forschungsfragen erschien es sinnvoll, pro Erklär- bzw. Begründungslevel entsprechend Punkte zu vergeben (s. o.): Ein Proband, dessen Antwort beispielsweise als E3 eingestuft wurde, erhielt somit 3 Punkte. Analog wurden auch bei den Begründungsaufgaben beispielsweise für eine B2-Antwort folglich 2 Punkte vergeben. Nicht bearbeitete Aufgaben und Falschantworten z. B. wurden entsprechend ihrem Level (E0 bzw. B0) mit 0 Punkten bewertet.

Bei den Ankreuzaufgaben wurde jeweils 1 Punkt vergeben, sofern diese korrekt beantwortet wurde. Bei manchen Aufgaben, bei denen zwei Kreuzchen notwendig waren, um als vollständig korrekt gewertet werden zu können (Aufgaben 2 f, 5 b, 6 b und 7 b des Testhefts), gab es 2 Punkte, wenn beide Kreuzchen korrekt gesetzt waren, und 1 Punkt, wenn ein Kreuzchen richtig gesetzt war. Widersprachen sich allerdings zwei angekreuzte Antworten inhaltlich, so wurde dieses Item mit 0 Punkten gewertet (vgl. Abb. 4-10).

Aufgabe 5 („Lesen und Rechtschreibung“)

- ☐ Wer mehr liest, macht auch durchschnittlich mehr Fehler.
- ✓ ☒ Je mehr Stunden ein Schüler durchschnittlich liest, desto weniger Fehler macht er im Test.
- ☐ Je größer die Fehleranzahl im Rechtschreibtest, desto geringer ist die durchschnittliche Lesezeit pro Monat.
- f ☒ Zwischen der Leistung im Rechtschreibtest und der Anzahl der Stunden, die ein Schüler pro Monat liest, besteht kein Zusammenhang.

Abb. 4-10: Auszug aus Aufgabe 5, bei der zwei widersprüchliche Aussagen angekreuzt wurden. Sie wird mit 0 Punkten gewertet.

Anmerkung: „Echte“ inhaltlich Widersprüche (Probanden erfassen durch Ankreuzen sowohl eine negative als auch eine positive Korrelation, z. B. durch Auswahl der Kästchen 1 und 2 bzw. 1 und 3) waren bei den 0-Punkte-Antworten mit einer Auftretensquote von < 1 % verschwindend gering, gleichwohl die Widersprüchlichkeit dieser Aussagen auch ohne Datenmaterial ersichtlich wäre.

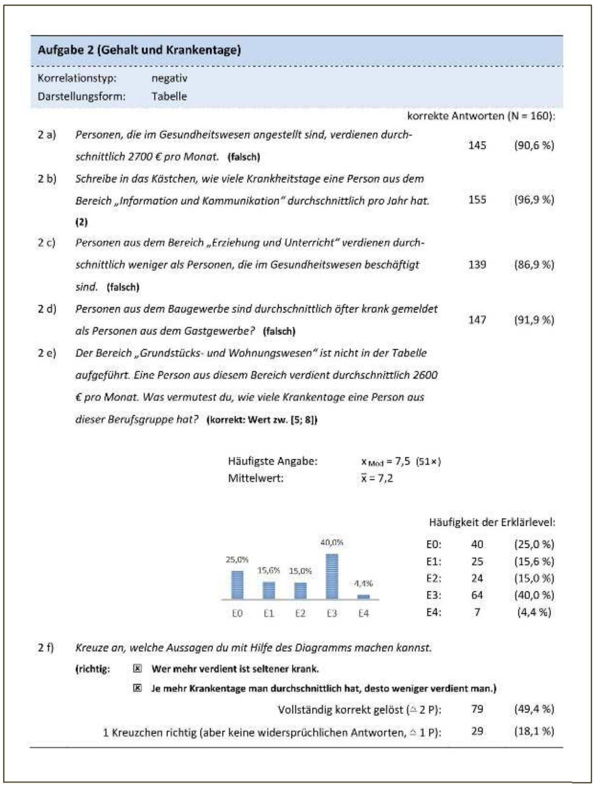
Für die übrigen 8,75 % aller 0-Punkte-Kandidaten bei diesen Ankreuzaufgaben (mit zwei richtigen Antwortmöglichkeiten) scheint allerdings das Erfassen einer Tendenz bzw. Korrelation in Verbindung mit der Aussage, zwischen zwei Merkmalen bestände kein Zusammenhang, offenbar kein Widerspruch zu sein.

5 Ergebnisse

Der Interpretation der Ergebnisse und der Beantwortung der Forschungsfragen ist eine rein deskriptive Auswertung vorangestellt. Sie liefert vorrangig die kodierten Daten und Werte, die unter anderem für eine spätere Umgruppierung und (Teil-) Zusammenfassung für die Beantwortung der Forschungsfragen die Grundlage bilden.

5.1 Deskriptive Auswertung der Fragebögen

Die Kodierung und Auswertung der 160 Fragebögen erfolgte gemäß den in Kap. 4.3 dargestellten Auswertungsprinzipien. Die detaillierten Werte können bei Bedarf dem Anhang (S. 146 ff) entnommen werden: Absolute und prozentuale Angaben über das Abschneiden der Schüler/innen pro Aufgabe inklusive der grundsätzlichen Information über den jeweiligen Korrelationstyp und die jeweilige Darstellungsform pro Aufgabe können dort ebenso nachgeschlagen werden wie die Werte, die als *optimal* festgesetzt (= rechnerisch ermittelt über eine Regressionsanalyse) wurden. Ebenso sind die ermittelten Modalwerte und das arithmetische Mittel zu den von den Schüler/inne/n abgegebenen Werten angeführt (vgl. Abb. 5-1).


Aufgabe 2 (Gehalt und Krankentage)

Korrelationstyp: negativ
Darstellungsform: Tabelle

korrekte Antworten (N = 160):

2 a)	*Personen, die im Gesundheitswesen angestellt sind, verdienen durchschnittlich 2700 € pro Monat.* **(falsch)**	145	(90,6 %)
2 b)	*Schreibe in das Kästchen, wie viele Krankheitstage eine Person aus dem Bereich „Information und Kommunikation" durchschnittlich pro Jahr hat.* **(2)**	155	(96,9 %)
2 c)	*Personen aus dem Bereich „Erziehung und Unterricht" verdienen durchschnittlich weniger als Personen, die im Gesundheitswesen beschäftigt sind.* **(falsch)**	139	(86,9 %)
2 d)	*Personen aus dem Baugewerbe sind durchschnittlich öfter krank gemeldet als Personen aus dem Gastgewerbe?* **(falsch)**	147	(91,9 %)
2 e)	*Der Bereich „Grundstücks- und Wohnungswesen" ist nicht in der Tabelle aufgeführt. Eine Person aus diesem Bereich verdient durchschnittlich 2600 € pro Monat. Was vermutest du, wie viele Krankentage eine Person aus dieser Berufsgruppe hat?* **(korrekt: Wert zw. [5; 8])**		

Häufigste Angabe: $x_{Mod} = 7,5$ (51×)
Mittelwert: $\bar{x} = 7,2$

Häufigkeit der Erklärlevel:

E0:	40	(25,0 %)
E1:	25	(15,6 %)
E2:	24	(15,0 %)
E3:	64	(40,0 %)
E4:	7	(4,4 %)

2 f) *Kreuze an, welche Aussagen du mit Hilfe des Diagramms machen kannst.*
(richtig: ☒ **Wer mehr verdient ist seltener krank.**
☒ **Je mehr Krankentage man durchschnittlich hat, desto weniger verdient man.)**

Vollständig korrekt gelöst (≙ 2 P):	79	(49,4 %)
1 Kreuzchen richtig (aber keine widersprüchlichen Antworten, ≙ 1 P):	29	(18,1 %)

Abb. 5-1: Beispiel einer deskriptiven Auswertung (hier Aufgabe 2)

Diese rein deskriptive Auswertung kann hilfreich und informativ sein, wenn man sich ggf. eine einzelne Aufgabe gezielt ansehen möchte, warum gewisse Aufgaben u. U. besonders gut oder z. B. auch sehr nachlässig bearbeitet wurden. Für die

Beantwortung der Forschungsfragen werden die Werte erst sinnhaft, wenn man sie nach der Fragestellung situationsgerecht gruppiert und mit weiteren (erhobenen) Daten in Kontext bringt. Dies soll nun im folgenden Kapitel geschehen.

5.2 Untersuchung der Forschungsfragen

Für einen besseren Überblick, welche Aufgaben aus dem Testheft jeweils zur Auswertung der Forschungsfragen herangezogen wurden, zeigt die Abbildung 5-2 vorab eine tabellarische Übersicht.

Forschungsfrage \ Aufgabe im Testheft	1a	1b	1c	1d	1e	2a	2b	2c	2d	2e	2f	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b
1a	x	x				x	x																				
1b			x	x				x	x																		
2a					x					x	x			x		x	x	x	x	x	x	x				x	
2b															x							x					x
2c					x					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2d																		x	x	x	x	x	x			x	x
2e					x					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	reading …		*reading between …*		*reading beyond … the data*	*reading …*		*reading between …*		*reading beyond … the data*																	

Abb. 5-2: Übersicht über die herangezogenen Items zur Untersuchung der Forschungsfragen

Dieser Tabelle lässt sich sogleich entnehmen, dass für die Beantwortung der ersten beiden Forschungsfragen 1 a und 1 b (S. 84 f) ausschließlich die ersten beiden *ausführlichen* Aufgaben 1 und 2 (vgl. Hinweise auf S. 97) des Testhefts herangezogen wurden, die somit nicht nur für die Motivation und Einführung in die Thematik gedacht waren, sondern eben zusätzlich Aussagen über die reine *Lesekompetenz* in und zwischen den Tabellenzeilen ermöglichen sollten (*reading the data* und *reading between the data*). Dass Schüler/innen dies sicher und zuverlässig in *Diagrammen* beherrschen, wurde im Kapitel 2.5 (Forschungsstand) bereits dargelegt. Die Auswertung der vorliegenden Studie kann nun ergänzen, dass sich diese Ergebnisse auch auf das Auslesen von Tabellen übertragen lassen. Die erste Forschungsfrage

F1 a) Wie erfolgreich entnehmen Schüler/innen einer (bivariaten) tabellarischen Darstellung konkret nachgefragte Werte?

kann im Rahmen dieser Untersuchung eindeutig äußerst positiv beantwortet werden: Über 96,9 % der Probanden lösten die Aufgaben 1 a, 1 b, 2 a und 2 b fehlerfrei (Abb. 5-3, S. 108). Die Erfolgsquote liegt dabei ähnlich hoch wie bei beispielsweise in den Studien von MORITZ (2004), ESTEPA und BATANERO (1996) oder PEREIRA-MENDOZA und MELLOR (1990), bei der knapp 98 % der Sechstklässler das Kompetenzlevel I (*reading the data*) erreichten.
Für die Beantwortung der eng mit der ersten Forschungsfrage in Verbindung stehenden zweiten Frage

F1 b) In welchem Ausmaß können Schüler/innen bei zweidimensionalen Tabellen Fragestellungen beantworten, die auf einen Vergleich zweier Merkmale zielen?

wurden ebenfalls ausschließlich die ersten beiden Aufgaben des Testhefts herangezogen (Aufgaben 1 c, 1 d, 2 c, 2 d). Dabei fällt das starke Absinken der Erfolgsquote beim Vergleichen von Datenpaaren innerhalb einer Tabelle (*reading between the data*) in der vorliegenden Untersuchung deutlich schwächer aus als bei vergleichbaren Studien mit Diagrammen, denn zwischen 85,6 % und 95,6 % lösten auch diese *Reading between the data*-Aufgaben korrekt (vgl. Abb. 5-3). Das bessere Abschneiden beim Vergleichen von Daten bzw. das Bewerten vergleichender Aussagen innerhalb einer Tabelle ist sicher auch auf die geringere Wahrscheinlichkeit von Ablesefehlern zurückzuführen, die PEREIRA und MELLOR in ihrer Studie erwähnen („*these errors are due to problems in the generalised reading/language or computation frames*", PEREIRA-MENDOZA & MELLOR 1990, p. 154), da die Daten bzw. Wertepaare *direkt* vorliegen und so leichter miteinander verglichen werden können. Allerdings muss – auch bei der späteren Einordnung neuer Werte – berücksichtigt werden, dass die Tabellen mit maximal 14 Wertepaaren recht überschaubar und bereits nach einem Merkmal vorsortiert angegeben waren, was die Orientierung innerhalb einer Tabelle in der Regel erheblich erleichtert.

Aufgabe 1 (Lebenserwartung und Herzfrequenz von Tieren)			
Korrelationstyp: negativ Darstellungsform: Tabelle			
		korrekte Antworten (N = 160):	
1 a)	*Das Herz eines Kaninchens schlägt durchschnittlich 215 Mal pro Minute.* **(richtig)**	**158**	**(98,8 %)**
1 b)	*Schreibe in das Kästchen, welche Lebenserwartung ein Fuchs durchschnittlich hat.* **(13)**	**158**	**(98,8 %)**
1 c)	*Ein Storch hat eine durchschnittlich höhere Lebenserwartung als ein Huhn.* **(richtig)**	**153**	**(95,6 %)**
1 d)	*Die durchschnittliche Herzfrequenz eines Meerschweinchens ist höher als die eines Goldhamsters.* **(falsch)**	**137**	**(85,6 %)**

Aufgabe 2 (Gehalt und Krankentage)			
Korrelationstyp: negativ Darstellungsform: Tabelle			
		korrekte Antworten (N = 160):	
2 a)	*Personen, die im Gesundheitswesen angestellt sind, verdienen durchschnittlich 2700 € pro Monat.* **(falsch)**	**145**	**(90,6 %)**
2 b)	*Schreibe in das Kästchen, wie viele Krankheitstage eine Person aus dem Bereich „Information und Kommunikation" durchschnittlich pro Jahr hat.* **(2)**	**155**	**(96,9 %)**
2 c)	*Personen aus dem Bereich „Erziehung und Unterricht" verdienen durchschnittlich weniger als Personen, die im Gesundheitswesen beschäftigt sind.* **(falsch)**	**139**	**(86,9 %)**
2 d)	*Personen aus dem Baugewerbe sind durchschnittlich öfter krank gemeldet als Personen aus dem Gastgewerbe?* **(falsch)**	**147**	**(91,9 %)**

Abb. 5-3: Tabelle mit der Anzahl der korrekten Antworten bei reinen „Auslese"-Aufgaben

Beinahe alle in Kapitel 2.5 angeführten Studien weisen darauf hin, dass Schüler/innen zum Teil erhebliche Schwierigkeiten haben, anhand des vorliegenden Datenmaterials Prognosen zu erstellen. Inwieweit sich die vorliegende Untersuchung mit diesen Ergebnissen kongruent zeigt, sollen v. a. die Forschungsfragen 2 a und 2 b zu klären versuchen.

F2 a) Wie schlüssig ordnen Schüler/innen neue Werte in ein bestehendes Datengefüge ein? Welche Qualität hat dabei die dafür angegebene Erklärung?

Nicht nur die bloße Einordnung von Werten, sondern auch der dahinterliegende Erkläransatz war in diesem Fall von besonderem Interesse. Dafür wurden die Fragen 1 e, 2 e, 4 a, 5 a, 6 a, 7 a, 8 a und 10 a (ausschließlich Aufgaben mit positiv bzw. negativ korrelierenden Werten) herangezogen, in denen je ein weiterer, noch nicht erfasster Wert in das bestehende Datengefüge einzuordnen war („Welchen Wert vermutest du ...?“). Bei jeder dieser Aufgaben wurde im Schnitt von 116 Studienteilnehmern (72,2 %) mindestens das Level E1 erreicht; das bedeutet, sie ordneten den Wert so ein, dass er innerhalb des tolerierten Intervalls lag und als korrekt gewertet werden konnte. Knapp jeder vierte Proband dieser Gruppe (25,2 %) war jedoch nicht in der Lage, eine sinnvolle oder nachvollziehbare Erklärung für den angegebenen (korrekten) Wert zu liefern (häufigste Level E1-Antwort: „Ich habe geschätzt.“).

	Aufgabe									
Level	1 e)	2 e)	4 a)	5 a)	6 a)	7 a)	8 a)	10 a)	$\bar{x}$	%
E0	40	40	62	43	29	38	45	59	(44,5)	(27,8 %)
E1	25	25	28	30	21	38	41	25	29,1	**18,2 %**
E2	27	24	26	27	16	28	21	10	22,4	**14,0 %**
E3	58	64	33	42	89	47	49	65	55,9	**34,9 %**
E4	10	7	11	18	5	9	4	1	8,13	**5,1 %**
∑	120 (160)	120 (160)	98 (160)	117 (160)	131 (160)	122 (160)	115 (160)	101 (160)		

Abb. 5-4: Tabelle mit den absoluten Häufigkeiten der auftretenden E-Stufen inkl. des arithmetischen Mittels

Rund 54 % aller Schüler/innen erreichten im Schnitt bei der Einordnung neuer Werte bei diesen Aufgaben mindestens das Erklärlevel E2, d. h. mehr als jeder zweite Untersuchungsteilnehmer konnte folglich erfolgreich den genannten Wert in das bestehende Datengefüge einordnen und dazu eine nachvollziehbare Erklärung angeben. Die Quote für das Erreichen des E2-Levels ist von Aufgabe zu Aufgabe unterschiedlich und bewegt sich bei allen Aufgaben zwischen 46,3 % und 59,4 %, sofern man die Aufgaben 4 a und 6 a außen vor lässt (s. u.). Die Tabelle in

Aufgabe		Level	abs. Häufigkeit (N = 160)		prozentual	
1 e)	*Eine Giraffe hat eine durchschnittliche Herzfrequenz von 75. Was vermutest du, wie alt so eine Giraffe wird?*	E0	40		25,0 %	
		E1	25		15,6 %	
		E2	27		16,9 %	
		E3	58	∑ 95	36,3 %	**∑ 59,4 %**
		E4	10		6,3 %	
2 e)	*Der Bereich „Grundstücks- und Wohnungswesen“ ist nicht in der Tabelle aufgeführt. Eine Person aus diesem Bereich verdient durchschnittlich 2600 € pro Monat. Was vermutest du, wie viele Krankentage eine Person aus dieser Berufsgruppe hat?*	E0	40		25,0 %	
		E1	25		15,6 %	
		E2	24		15,0 %	
		E3	64	∑ 95	40,0 %	**∑ 59,4 %**
		E4	7		4,4 %	
4 a)	*Angenommen, die kleine Forschungsgruppe führt eine weitere Schallpegelmessung in einer Klasse mit 22 Schülern durch. Welchen Wert würden sie deiner Meinung nach erhalten?*	E0	62		38,8 %	
		E1	28		17,5 %	
		E2	26		16,3 %	
		E3	33	∑ 70	20,6 %	**∑ 43,8 %**
		E4	11		6,9 %	
5 a)	*Ein anderer Schüler, der bei diesem Test nicht mitgeschrieben gibt an, dass er ca. 11 Stunden pro Monat liest. Was glaubst du, wie viele Rechtschreibfehler er wohl haben würde, wenn er auch diesen Test schreibt?*	E0	43		26,9 %	
		E1	30		18,8 %	
		E2	27		16,9 %	
		E3	42	∑ 87	6,3 %	**∑ 54,4 %**
		E4	18		11,3 %	
6 a)	*Eine weitere Testperson im Alter von 42 Jahren macht denselben Reaktionstest. Was schätzt du, wie viele Sekunden diese Person durchschnittlich braucht, um den richtigen Knopf zu drücken?*	E0	29		18,1 %	
		E1	21		13,1 %	
		E2	16		10,0 %	
		E3	89	∑ 110	55,6 %	**∑ 68,8 %**
		E4	5		3,1 %	
7 a)	*An einem sonnigen Tag hat es mittags 14° C. Was denkst du, wie viele Verkehrskontrollen auf Herrn Müllers Handy-App angezeigt werden?*	E0	38		23,8 %	
		E1	38		23,8 %	
		E2	28		17,5 %	
		E3	47	∑ 84	29,4 %	**∑ 52,5 %**
		E4	9		5,6 %	
8 a)	*Offenbar wurde kein Elfjähriger zu diesem Thema befragt. Was würdest du vermuten, wie viel Taschengeld er pro Woche bekommt?*	E0	45		28,1 %	
		E1	41		25,6 %	
		E2	21		13,1 %	
		E3	49	∑ 74	30,6 %	**∑ 46,3 %**
		E4	4		2,5 %	
10 a)	*Im Jahr 2002 verspeiste der deutsche Durchschnittsbürger 34,5 kg Tiefkühlkost. Welchen Stromverbrauch vermutest du in diesem Jahr in einem deutschen Durchschnittshaushalt?*	E0	59		36,9 %	
		E1	25		15,6 %	
		E2	10		6,3 %	
		E3	65	∑ 76	40,6 %	**∑ 47,5 %**
		E4	1		0,6 %	

Abb. 5-5: Tabelle mit den detaillierten Werten jeder E-Stufe in absoluten Zahlen und in Prozent.

Abbildung 5-5 zeigt die Häufigkeit des jeweils erreichten Levels je nach Aufgabe. Die hohe Erfolgsquote bei Aufgabe 6 a („Wer ist der Schnellste?“; 68,8 % der Probanden erreichte hier durchschnittlich mindestens das Level E2) lässt sich sicherlich dadurch erklären, dass man bei dem einzuordnenden neuen Wert relativ leicht das arithmetische Mittel der benachbarten Werte (0,80 s und 0,90 s) ermitteln konnte (tatsächlich war die häufigste Angabe hier $x_{Mod} = 0{,}85$ mit 86 Nennungen), während bei allen anderen Aufgaben die bloße Angabe des arithmetischen Mittelwertes nicht ohne weiteres möglich bzw. sinnvoll erschien (z. B. 7,5 Krankentage oder 5,5 Rechtschreibfehler).

Die im Vergleich relativ niedrige Erfolgsquote bei Aufgabe 4 a dagegen ist wohl dem Umstand geschuldet, dass die persönliche Erwartungshaltung (die erwartete Korrelation) der dargestellten Korrelation zuwider läuft (vgl. die Ergebnisse der Studien von MORITZ (2004), ZIEFFLER (2008) und ESTEPA & BATANERO): Nicht wenige Schüler/innen erklärten in der sich anschließenden Teilaufgabe 4 b beispielsweise: *„Ich finde, es ist genau das Gegenteil: Je mehr Schüler, desto lauter ist es im Klassenzimmer."* bzw. *„Das kann man nicht genau sagen, weil es ja von den Schülern abhängt, wie laut die gerade sind."*

Bemerkenswert ist, dass bei über 90 % der abgegebenen Antworten (626 von 691), die mindestens das Level E2 oder höher aufweisen, eine rein *lokale* Argumentation zugrunde liegt, d. h. für die (korrekte) Einordnung des vorgegebenen Wertes gründet sich der Erklärungsansatz des Probanden lediglich auf das Heranziehen direkt benachbarter Wertepaare (= Level E2 und E3). Diese Vorgehensweise scheint praktikabel und legitim, lässt aber theoretisch nicht zwangsläufig den Schluss zu, dass Schüler/innen den Trend in dem vorliegenden Datensatz notwendigerweise erkannt haben. Allerdings könnte man argumentieren, dass eine erfolgreiche Einordnung neuer Daten durchaus darauf schließen lässt, dass sehr wohl ein Trend erfasst wurde, auch wenn sich ggf. der Proband dessen nicht explizit bewusst war: Die grafische Aufbereitung bzw. Vorsortierung der Daten, der zugrundeliegende hohe Korrelationskoeffizient sowie eine der Testsituation geschuldete a-priori-Annahme eines vorhandenen Trends spielen hier sicher eine Rolle.

Die Annahme, dass einer sinnvollen Einordnung eines weiteren Wertes das implizite Erfassen eines Trends zugrunde liegt, lässt sich bestärken, indem man prüft, inwieweit Probanden der Stufe E2 und höher auch die nachfolgende Ankreuzaufgabe widerspruchsfrei gelöst haben (d. h. sie haben mindestens eine korrekte Aussage zum Trend angekreuzt und keine widersprüchlichen Angaben gemacht).

Aufgabe	Anzahl der Probanden mit mind. Stufe E2	Anzahl der Probanden, die zugleich mind. 1 P. bei der nachfolgenden Ankreuz-aufgabe erzielten	Prozent	
2 e & 2 f	95	70	73,7 %	∅ 74,0 %
5 a & 5 b	87	67	77,0 %	
6 a & 6 b	110	84	76,4 %	
7 a & 7 b	84	58	69,0 %	

Abb. 5-6: Erfolgsquote bei der Einordnung mit anschließender Ankreuzaufgabe

Rund 74 % der Studienteilnehmer (Abb. 5-6) beantworteten nach einer korrekten Einordnung des vorgegebenen Werts auch die Ankreuzaufgabe richtig, was als Indiz für das erfolgreiche Erfassen des zugrundeliegenden Trends gewertet werden kann.

Die Fähigkeit, neue, noch nicht dargestellte Werte in ein bestehendes Datengefüge einzuordnen, könnte man zusammenfassend also durchaus als zufriedenstellend oder wenigstens als ausreichend bezeichnen; zumindest scheint es nicht gerechtfertigt, die Performance als *„extremely low"* (PEREIRA-MENDOZA & MELLOR 1990, p. 154) zu bezeichnen. Es ist anzunehmen, dass der betragsmäßig hohe Korrelationsfaktor in diesen Aufgaben (vgl. die Forschungsergebnisse von ESTEPA & BATANERO) sowie die Vorsortierung der Tabellen die Einordnung neuer Werte vermutlich insoweit vereinfacht haben, so dass diese im Vergleich zu den bisherigen Studien relativ hohe Erfolgsquote möglich wurde.

Eng verbunden mit dieser *Einordnungskompetenz* ist die bereits angedachte Frage, wie Schüler/innen den vermeintlich erfassten Trend in den Daten *begründen*.

F2 b) Inwieweit korreliert die Fähigkeit, neue Werte schlüssig in ein bestehendes Datengefüge einzuordnen, mit der Fähigkeit, einen Trend in Daten sachgerecht zu begründen?

Für die Beantwortung dieser Forschungsfrage wurden die Aufgaben 4, 8 und 10 herangezogen, denn hier sollten die Schüler/innen nach einer Einordnung eines neuen Werts in das vorgegebene Datengefüge anschließend entweder mit eigenen Worten begründen, ob und warum sie in dem vorliegenden Datensatz einen Zusammenhang zwischen den Merkmalen erkennen („Erkennst du einen Zusammenhang ...? Begründe deine Antwort." (A 10)), oder zu einer bereits formulierten Behauptung bzgl. eines Trends Stellung nehmen („Max/Jana behauptet ... Was meinst du dazu? Begründe deine Antwort ..." (A 4, A 8).

Für einen ersten Vergleich wurde jeweils geprüft, inwieweit das Erklärlevel und das Begründungslevel – sie wurden beide, wie in Kap. 4.3 dargestellt, bewusst in je fünf Abstufungen „hierarchisch parallel" konzipiert (vgl. S. 101) – bei diesen drei Aufgaben jeweils korrelieren (Abb. 5-7).

Aufgabe 4 a + b (Lärm im Klassenzimmer):	$r_S \approx 0{,}24$
Aufgabe 8 a + b (Mehr Taschengeld!):	$r_S \approx 0{,}39$
Aufgabe 10 a + b (Tiefkühlkost und Stromverbrauch):	$r_S \approx 0{,}36$

(r_s ist der Rangkorrelationskoeffizient nach Spearman)

Abb. 5-7: Korrelation zwischen Erklär- und Begründungslevel

Bei allen drei Aufgaben zeigt sich zwischen Erklär- und Begründungslevel eine eher geringe Korrelation; das heißt, Probanden, die durchschnittlich einen hohen (bzw. niedrigen) Score beim Einordnen von Werten inklusive entsprechender Erklärung erreichten (Erklärlevel), schnitten bei den Beurteilungs- und Begründungsaufgaben (Begründungslevel) deswegen nicht automatisch ähnlich gut (bzw. wenig zufriedenstellend) ab. Die unter Umständen unausgesprochene vorschnelle Annahme „Wer gut im Einordnen von Werten ist, ist bestimmt auch gut im Beurteilen und Begründen statistischer Trends" darf deutlich zurückgewiesen werden.

Vergleicht man in einem nächsten Schritt die E- und B-Level anhand der arithmetischen Mittelwerte, wird ersichtlich, dass die Schüler/innen bei jeder dieser Aufgaben ein durchschnittlich deutlich geringeres Begründungslevel als Erklärlevel aufweisen (Abb. 5-8). Da die Klassifizierung der Antworten sowohl bei den Erklär- als auch bei den Begründungsaufgaben in je fünf Abstufungen vorgenommen wird und entsprechend Punkte vergeben werden (erreichtes Level $\Rightarrow$ dementsprechende Punktzahl), sind die arithmetischen Mittelwerte hier eine geeignete Kenngröße für einen Vergleich dieser Art.

	Aufgabe 4	Aufgabe 8	Aufgabe 10
Durchschnittliches E-Level	1,38	1,53	1,52
	↓	↓	↓
Durchschnittliches B-Level	1,03	1,11	0,96

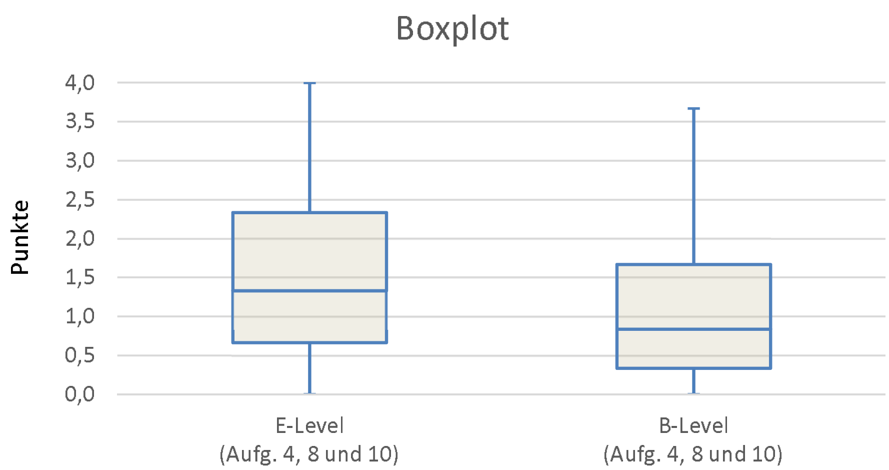


Abb. 5-8: Durchschnittlich erreichte Punktzahl bei Aufgaben mit Einordnung/Erklärung und anschließender Begründung

Ein genauerer Blick auf die Auswertungstabelle, z. B. auf die besten 20 Probanden hinsichtlich des E-Levels bei diesen drei Aufgaben (im dargestellten Fall sortiert nach der Spalte „∅“ der arithmetischen Mittelwerte der jeweiligen E-Level, vgl. Abb. 5-9 auf S. 115), zeigt, dass trotz durchschnittlich hohem E-Level die Erfolgsquote beim Begründen deutlich abnimmt.

Probanden-nr.	E-Level 4a)	8a)	10a)	∅	B-Level 4b)	8b)	10b)	∅
30	0	4	3	2,33	0	1	1	0,67
78	2	3	3	2,67	0	1	1	0,67
60	2	3	2	2,33	1	1	4	2,00
62	4	3	0	2,33	3	2	4	3,00
69	2	3	3	2,67	3	1	4	2,67
122	0	1	2	1,00	0	0	0	0,00
58	2	3	3	2,67	3	3	4	3,33
56	2	3	3	2,67	1	2	0	1,00
81	3	3	3	3,00	4	0	0	1,33
75	3	3	3	3,00	0	0	3	1,00
126	3	3	3	3,00	0	0	4	1,33
55	3	3	3	3,00	2	1	1	1,33
45	1	4	3	2,67	0	1	1	0,67
71	3	3	3	3,00	1	0	4	1,67
98	3	3	3	3,00	0	3	0	1,00
102	4	4	3	3,67	4	4	0	2,67
109	3	3	3	3,00	4	3	4	3,67
18	3	3	3	3,00	0	0	4	1,33
73	4	2	3	3,00	0	2	4	2,00
121	4	4	4	4,00	2	1	0	1,00
	1,38	1,53	1,52		1,03	1,11	0,96	

Abb. 5-9: Auszug aus der Auswertungstabelle (Excel) mit Blick auf die 20 besten Studienteilnehmer bzgl. des E-Levels (Spalten „∅“).

Auch der diesbezüglich beste Studienteilnehmer (Nr. 121), der bei diesen drei Aufgaben jeweils das höchste E-Level erreichte, schaffte bei den Begründungsaufgaben beispielsweise durchschnittlich nur das Begründungslevel 1 (vgl. Abb. 5-9). Auch bei anderen relativ erfolgreichen E-Level-Kandidaten zeigte sich, dass das Begründungslevel nicht mit dem Erklärlevel Schritt halten kann.

Bemerkenswert: Unabhängig vom erreichten E-Level schafften 74,6 % der Probanden lediglich das B-Level 0 oder 1. Ein „pauschaler“, „konkreter“ oder gar „umfassender Datenbezug“ (≙ Level B2, B3 und B4, vgl. S. 102 f) konnte durchschnittlich nur knapp 41 der insgesamt 160 Studienteilnehmer attestiert werden.

Für diese defizitären Begründungsmuster scheinen hauptsächlich die folgenden beiden Gründe verantwortlich zu sein:

1. Zweifelsohne erfassen Schüler/innen durchaus den vorliegenden Trend in den Daten und ordnen dementsprechend neue Werte in das bestehende Datengefüge korrekt ein. Allerdings haben sie offensichtlich wenig Übung darin, *wie* man das begründen kann. Selbst bei prinzipiell korrekten Begründungen bleiben die Ausführungen der Probanden häufig zu oberflächlich und pauschal formuliert:

„Das könnte zutreffen, ist aber nicht immer so." (≙ B1, Proband #121)

„Das Diagramm sagt das ... und somit gebe ich Max recht." (≙ B2, Proband #55).

„Ja das stimmt, aber irgendwie ist das unlogisch." (≙ B1, Proband #10)

Es mangelt den Schüler/inne/n anscheinend an der notwendigen Argumentationstechnik, um hier kompetent auftreten zu können. Das Beschreiben von Entwicklungen unter Zuhilfenahme abgebildeter Wertepaare, die Benennung der Merkmalsausprägungen oder die Berücksichtigung der Achsen bzw. deren Skalierung wären wichtige Komponenten einer sachgerechten Begründung, auf die die Schüler/innen allerdings nicht zurückgreifen bzw. nicht zurückgreifen können. Oder um es mit den Formulierungen der KMK-Standards zu beschreiben: Die prozessbezogenen mathematischen Kompetenzen „mathematisch Argumentieren" (K1) und „kommunizieren" (K6) scheinen bei Schüler/inne/n der Sekundarstufe I nicht adäquat ausgeprägt zu sein.

2. Ein weiterer entscheidender Faktor scheint die subjektive Überlagerung der gegebenen Antworten zu sein. Selbst bei Schüler/inne/n mit einem durchschnittlich hohen E-Level sinkt das entsprechende B-Level deutlich ab, da sie in ihren Begründungen rein mit ihrer subjektiven Sichtweise bzw. ihrem persönlichen Erfahrungshorizont antworten; die vorliegenden Daten lassen sie dabei völlig außer Acht:
   - *„Das trifft nicht ganz zu, da es die Eltern bestimmen, wie viel sie jedem Kind geben."* (Proband #106, hier Aufgabe 8)
   - *„Nein, da es auch noch andere Geräte gibt, die Strom verbrauchen"* (Proband #114, Aufgabe 10)
   - *„Das kann eigentlich gar nicht sein, da mehr Personen auch mehr Lärm machen."* (Proband #73, Aufgabe 4) oder
   - *„Das ist Quatsch. Denn eine Klasse mit 20 Schülern kann genauso laut sein wie eine Klasse mit 30 Schülern."* (Proband #18, Aufgabe 4)

Allein bei Aufgabe 8 (Taschengeld) wiesen 97 der 160 Antworten (60,6 %) eine subjektive Färbung in der Begründung auf. Selbst der explizite Hinweis „Begründe

mit Hilfe des Diagramms“ wurde in fast allen diesen Fällen ignoriert. Schüler/innen antworten also nicht selten „aus dem Bauch heraus“ und nicht mithilfe des vorliegenden Datenmaterials. Auch dieser Aspekt spricht für eine gering ausgeprägte Fähigkeit, mathematisch zu argumentieren (K1).
Zusammenfassend lässt sich die Forschungsfrage 2 b dahingehend beantworten, dass Schüler/innen, die neue Werte in ein bestehendes Datengefüge erfolgreich einordnen können, mit hoher Wahrscheinlichkeit auch die dahinterliegende Korrelation erfassen (vgl. F2 a), allerdings gelingt es ihnen äußerst selten, diese Korrelation mit eigenen Worten nachvollziehbar und mathematisch korrekt argumentierend zu begründen.

Besonders im Falle des Nichterkennens einer Korrelation oder bei Aufgaben ohne korrelierende Merkmale (Aufgabe 3 und 9) ist zudem beobachtbar, dass für Schüler/innen ein Nein keiner weiteren Erklärung bedarf: Ist in einem Datensatz kein Trend zu erkennen, ist dieses „Nein“ häufig eine Feststellung, die in ihren Augen anscheinend nicht weiter begründet werden muss. Auch hier fehlt offenbar das nötige Rüstzeug für eine sachgerechte Argumentation, denn zumindest ein Eingehen auf die Streuung der Datenpunkte o. ä. wäre der Minimalkonsens für eine aus mathematischer Sicht akzeptable Antwort.

F2 c) Inwiefern ist das Einordnen, Erklären und Begründen von Zusammenhängen abhängig vom Alter der Probanden?

Beinahe alle im Kapitel 2.5 dargestellten Untersuchungen legen dar, dass sich ältere Schüler/innen deutlich datenkompetenter zeigen als jüngere. Für eine eventuelle Verifikation dieser Tendenz wurden in der vorliegenden Untersuchung die Punkte aller E- und B-Level sowie die erreichte Punktzahl in allen dazugehörigen Ankreuzaufgaben (= alle Aufgaben außer die reinen *Ableseaufgaben* 1 a - 1 d und 2 a - 2 d) bei jedem Probanden addiert, so dass diese Summe – ähnlich wie bei typischen schriftlichen Leistungsnachweisen im Unterrichtsalltag – die Möglichkeit eröffnet, die Leistungen untereinander zu vergleichen.

Aufgabe	∅ Punkte 6.-Klässler (N = 83)	∅ Punkte 9.-Klässler (N = 77)	(max. erreichbare Punktzahl)
1 e)	1,70	1,97	(4)
2 e)	1,75	1,92	(4)
2 f)	1,25	1,08	(2)
3 a)	0,53	0,49	(1)
3 b)	0,82	0,86	(1)
4 a)	1,43	1,35	(4)
4 b)	0,94	1,13	(4)
5 a)	1,81	1,71	(4)
5 b)	1,30	1,01	(2)
6 a)	2,13	2,12	(4)
6 b)	1,10	1,10	(2)
7 a)	1,60	1,79	(4)
7 b)	0,90	0,82	(2)
8 a)	1,57	1,51	(4)
8 b)	1,24	0,96	(4)
9 a)	0,30	0,51	(1)
9 b)	0,82	0,91	(4)
10 a)	1,51	1,55	(4)
10 b)	1,47	1,96	(4)
$\bar{x}$	**24,17**	**24,75**	∑ 59
σ	**8,80**	**11,61**	
P_{max}	45	45	
x_{Mod}	20	29	

Legende: $\bar{x}$ *Arithmetisches Mittel*
σ Standardabweichung
P_{max} *Maximal erreichte Punktzahl*
x_{Mod} *Modalwert*

Abb. 5-10: Vergleich der Fähigkeiten bei 6.- und 9.-Klässlern

Entgegen den Ergebnissen bisheriger Studien zeigt sich deutlich, dass die Schüler/innen der 9. Jahrgangsstufe in dieser Untersuchung nicht signifikant besser abschneiden als diejenigen der 6. Jahrgangsstufe. Die arithmetischen Mittelwerte sind nahezu identisch, die höchste erreichte Punktzahl sogar exakt dieselbe; bei den 9.-Klässlern zeigt sich allerdings bei der Standardabweichung ein etwas höherer Wert. Ältere Schüler/innen erweisen sich hier also nicht datenkompetenter als ihre jüngeren Mitstreiter. Man könnte folglich verlautbaren, dass sich der Kompetenzzuwachs in der Sekundarstufe von Klasse 6 bis Klasse 9 nicht weiter ausbildet (was u. a. sicher daran liegt, dass dieser Themenbereich kein expliziter Bestandteil der Bildungspläne in der Sekundarstufe I ist).

F2 d) Hängt das Erkennen eines Zusammenhangs von der Art der Korrelation ab? Sind Schüler/innen beim Einordnen neuer Werte bei einer zugrundeliegenden positiven Korrelation erfolgreicher als bei einer negativen Korrelation?

Für einen Vergleich, ob Schüler/innen Aufgaben mit dahinterliegender positiver Korrelation erfolgreicher bearbeiten als Aufgaben mit negativer Korrelation, wurde einmal die erreichte Punktzahl aller Testteilnehmer bei allen Aufgaben mit positiver Korrelation addiert und ebenso die erreichte Punktzahl bei allen Aufgaben mit negativer Korrelation (Abb. 5-11).

	Aufgaben mit pos. Korrelation								Aufgaben mit neg. Korrelation						
Aufgabe	6a	6b	7a	7b	8a	8b	10a	10b	1e	2e	2f	4a	4b	5a	5b
Typ	E	☒	E	☒	E	B	E	B	E	E	☒	E	B	E	☒
∅	2,13	1,10	1,69	0,86	1,54	1,11	1,53	0,95	1,08	1,08	1,17	1,39	1,03	1,76	1,16
	∅ 1,36								∅ 1,24						
σ	0,76								0,65						

Abb. 5-11: Vergleich: Erkennen positiver vs. negativer Korrelationen

Vergleicht man die durchschnittlich erreichte Punktzahl der Aufgaben mit positiver Korrelation mit den Aufgaben mit negativer Korrelation, so lässt dies den Schluss zu, dass Schüler/innen bezüglich ihrer *Einordnungskompetenz* keinen Korrelationstyp präferieren; der errechnete Unterschied ist nicht signifikant. In der Literatur finden sich Hinweise, dass Schüler/innen positive Trends leichter erkennen als negative, jedoch relativieren die Autoren dieser Studien ihre Aussagen teilweise selbst, indem sie zu bedenken geben, dass offenbar nicht (nur) der Korrelationstyp, sondern auch die Stärke der Korrelation eine entscheidende Rolle spielt. In der Untersuchung von ESTEPA und BATANERO (1996) beispielsweise schnitten die Probanden bei der Aufgabe mit zugrundeliegender negativer Korrelation beinahe vier Mal besser ab (84,5 %) als bei der Aufgabe mit positiver Korrelation (21,6 %), allerdings weisen auch diese beiden Autoren darauf hin, dass die Erkennungsrate einer Tendenz von der Güte bzw. der Höhe des Korrelationskoeffizienten abhängig zu sein scheint (vgl. S 78 ff). Da in der vorliegenden Untersuchung alle dargebotenen Aufgaben einen Korrelationskoeffizienten $|\rho| > 0{,}9$ aufweisen, kann behauptet werden, dass es für die Schüler/innen bei der Einordnung neuer Werte in ein bestehendes Datengefüge wohl keine Rolle spielt, ob sich die zugrundeliegende Korrelation positiv oder negativ darstellt.

F2 e) Ist das erfolgreiche Einordnen neuer Werte in ein bestehendes Datengefüge abhängig von der Darstellungsform (Graph vs. Tabelle)?

Auch hier wurde pro Untersuchungsteilnehmer jeweils die erreichte Punktzahl aller Aufgaben addiert, gruppiert nach der jeweiligen Darstellungsform der Daten (vgl. Abb 5-12). Hier zeigt sich ebenfalls, dass es offensichtlich keinen Unterschied macht, ob bei der Einordnung neuer Werte das ursprüngliche Datenmaterial in graphischer oder in tabellarischer Form vorliegt. Da das Auslesen von Werten aus

	Aufgaben mit Diagramm											
Aufgabe	3a	3b	4a	4b	5a	5b	7a	7b	8a	8b	9a	9b
Typ	☒	☒	E	B	E	☒	E	☒	E	B	☒	B
∅	0,51	0,84	1,39	1,03	1,76	1,16	1,69	0,86	1,54	1,11	0,40	0,86
	∅ 1,10											
σ	0,42											

	Aufgaben mit Tabelle						
Aufgabe	1e	2e	2f	6a	6b	10a	10b
Typ	E	E	☒	E	☒	E	B
∅	1,08	1,08	1,17	2,13	1,10	1,53	0,95
	∅ 1,29						
σ	0,38						

Typ ≙ Aufgabentyp: ☒ = Ankreuzaufgabe, E = Erkläraufgabe, B = Begründungsaufgabe
∅ = arithm. Mittel,
σ = Standardabweichung

Abb. 5-12: Vergleich der Erfolgsquote bei Diagrammen und Tabellen

Tabellen und Diagrammen sicher beherrscht wird (vgl. Forschungsfrage 1 a + 1 b) und die Probanden in der Regel auf ein rein lokales Begründungsmuster zurückgreifen (das bloße Heranziehen benachbarter Punkte, vgl. Forschungsfrage 2 a), scheint es nachvollziehbar, dass die Erfolgsquote bei der Bearbeitung der

Diagramm- und Tabellenaufgaben ähnlich hoch bzw. nicht signifikant unterschiedlich ist.

Zudem wird die bereits erwähnte hohe Korrelation der Werte bei allen Aufgaben auch hier mitverantwortlich sein, dass die Schüler/innen die Aufgaben unabhängig von der dargebotenen Darstellungsform ähnlich erfolgreich bearbeiteten.

6 Zusammenfassung und Einordnung

Wenn man versuchte, die Ergebnisse dieser Studie in einem Absatz zusammenzufassen, würde dieser wohl wie folgt lauten:

> *Schüler/innen der Sekundarstufe I können unabhängig von ihrem Alter sowohl tabellarische als auch grafische Darstellungen bivariater Datensätze lesen und in das vorgegebene Datengefüge auch neue, noch nicht erfasste Werte sinnvoll einordnen. Dabei zeigen sie in den allermeisten Fällen ein lokales Erklärmuster, d. h. sie orientieren sich bei der Einordnung an den direkt benachbarten Wertepaaren. In der Regel erfassen sie prinzipiell auch den dargestellten Trend, allerdings haben sie deutliche Schwierigkeiten, ihre Erkenntnis sachgerecht zu formulieren bzw. zu begründen.*

Die vorliegende Studie liefert somit durchaus interessante Antworten auf die gestellten Forschungsfragen, auch wenn diese Erkenntnisse lediglich Facetten einer allgemeinen Datenkompetenz widerspiegeln. Datenkompetenz im Allgemeinen bei Schüler/inne/n in einer (einzigen) empirischen Studie erfassen zu wollen, würde allerdings ein nahezu unmögliches Unterfangen darstellen: Ein Vorhaben dieser Art würde sowohl bereits bei dem Versuch scheitern, Datenkompetenz eindeutig definieren zu wollen (vgl. Kap. 2.1), als auch an der Schwierigkeit, zu erfassen, welche Einzel- bzw. Teilkomponenten bzw. -kompetenzen dafür eine Rolle spielen (die sich zudem über Jahre aufbauen und kaum geklärt werden kann, wie diese sich wechselseitig beeinflussen). Dennoch sei hier abschließend der Versuch unternommen, zumindest auf der Grundlage der gewonnenen Erkenntnisse, Aussagen über die *bivariate Datenkompetenz* von 6.- bzw. 9.-Klässlern zu treffen.

Für eine erste Bewertung und Einordnung eines Testergebnisses wird dazu die Summe der erreichten Punkte aus allen zehn Aufgaben herangezogen. Die Kodierung und Zuordnung der Antworten zu einer bestimmten Stufe entsprechend ihrer Qualität und der daraus resultierenden *Vergütung* mit derselben Punktzahl ermöglichen diese Vorgehensweise.

Der Boxplot (Abb. 6-1) zeigt dabei die Verteilung der erreichten Punktezahlen aller 160 Probanden: 50 % der Studienteilnehmer erreichten ein Ergebnis zwischen 24 und 39 (von insgesamt 67 möglichen) Punkten, der Median lag bei 34 Punkten.

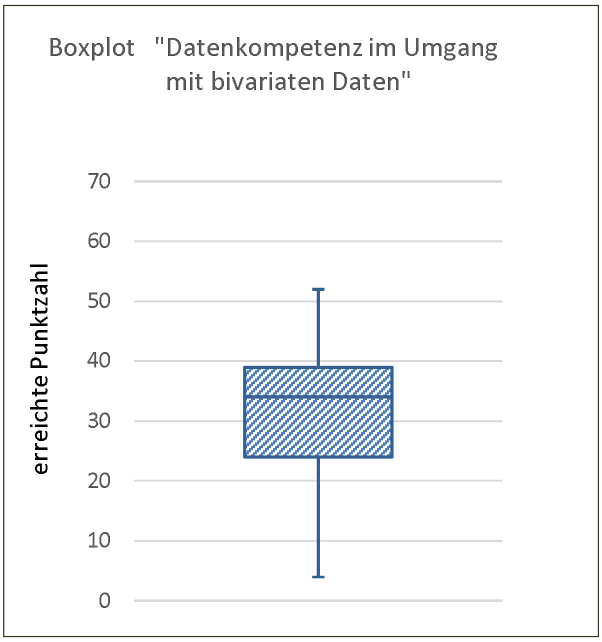


Abb. 6-1: Verteilung der erreichten Punktezahlen

Das reine Aufsummieren und Vergleichen von Punktzahlen stellt allerdings eine nur wenig differenzierte und bedingt aussagekräftige Möglichkeit dar, Schüler/innen eine gewisse Kompetenzklasse oder -stufe zuzuweisen. Ähnlich wie bei der Beurteilung schriftlicher Leistungsnachweise im Unterricht ermöglichen Punktezahlen zwar einen Vergleich der Testteilnehmer untereinander, differenzieren aber zu wenig die Stärken und Schwächen in den einzelnen Teilbereichen (eine Schülerin kann im Fach Englisch beispielsweise „sehr gut" im Prüfungsteil Vokabeln sein, im Grammatikteil aber nur ein „mangelhaft" erreichen; die Gesamtnote wäre „befriedigend").

Welche Rückschlüsse lässt nun eine bestimmte Punktzahl auf die gewählte Definition von (bivariater) Datenkompetenz zu? Welches Erklär- oder Begründungslevel könnte man bei einem Studienteilnehmer daraus ableiten?

Um eine Aussagemöglichkeit bezüglich des zu erwarteten E-Levels abhängig vom Gesmatscore zu erhalten, wird für die folgende Analyse die jeweils erreichte Gesamtpunktzahl mit dem durchschnittlich erreichten E-Level (= arithmetisches Mittel der erreichten Punkte in den herangezogenen acht Erkläraufgaben) in Zusammenhang gebracht. Das Streudiagramm dazu (Abb. 6-2) zeigt sogleich einen deutlichen Zusammenhang; der Rangkorrelationskoeffizient nach Spearman beträgt $\rho = 0{,}93$.

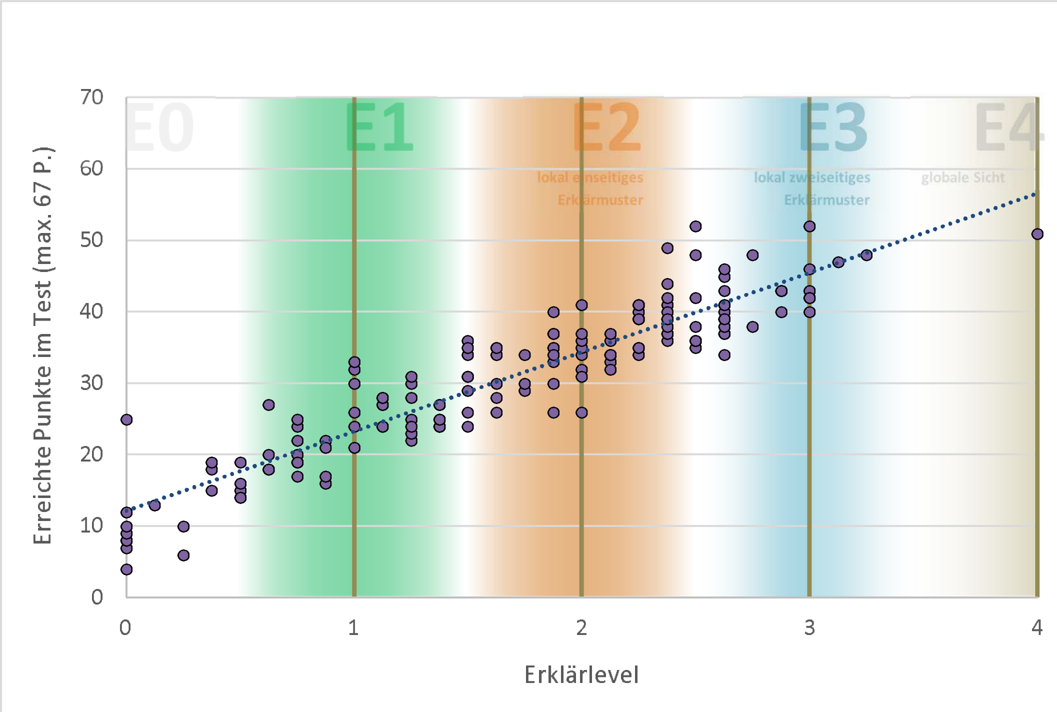


Abb. 6-2: Der offensichtliche Zusammenhang zwischen erreichter Gesamtpunktzahl und durchschnittlich erreichtem E-Level als Streudiagramm

Diese Korrelation ist bis zu einem gewissen Grad auch durchaus zu erwarten, da sich die Gesamtpunktzahl vorrangig aus den vergebenen Punkten bei den Erkläraufgaben errechnet; dennoch zeigt der hohe Korrelationswert ρ, dass sich auch die Ergebnisse der zugehörigen Ankreuzaufgaben sowie der (wenngleich deutlich weniger erfolgreich bewältigten) Begründungsaufgaben stimmig (= korrelierend) in das Gesamtbild einfügen.

Selbstverständlich lässt sich zum obigen Streudiagramm – wie im theoretischen Teil dargestellt (vgl. Kap. 2.2.7) – eine Regressionsgerade angeben, mithilfe derer man zu einer vorgegeben Punktzahl (y) das entsprechnde E-Level (x) errechnen kann.

Beispiel: Testergebnis: 34 Punkte

Regressionsgleichung (hier ermittelt mit Excel®): $y = 11{,}11x + 12{,}11$

$\Rightarrow 34 = 11{,}11x + 12{,}11 \Leftrightarrow x \approx 1{,}97$

$x \approx 2$ (≙ Erklärlevel 2)

Für die obige Beispielrechnung bedeutet das: Man sollte von einem beliebigen Testteilnehmer mit insgesamt 34 Punkten durchschnittlich das Erklärlevel 2 (= „lokal einseitig“, vgl. S. 100) oder höher erwarten können.

Selbstverständlich ließe sich dieses Procedere auch für das Zusammenspiel von erreichter Gesamtpunktzahl und Begründungslevel durchführen, allerdings scheint aufgrund der dürftigen Ergebnisse bei den Begründungsaufgaben (vgl. Abb. 6-3) diese Zuordnung kaum eine lohnenswerte Aussage hervorzubringen.

Durchschnittlich erreichtes B-Level: (ermittelt über die Aufgaben 4, 8, 9 und 10)

Arithmetisches Mittel:	$\bar{x} = 0{,}99$
Standardabweichung:	$\sigma = 0{,}8$
Zentralwert:	$x_{med} = 0{,}88$
Häufigster Wert:	$x_{Mod} = 0$ (24×)

86 % aller Probanden erreichten durchschnittlich nur das Level B0 bzw B1.

Abb. 6-4: Zusammenfassung über das wenig erfolgreiche Abschneiden bei den Begründungsaufgaben anhand der gängigen statistischen Kenngrößen

Die soeben noch als „nicht lohnenswert“ bezeichnete Darstellung von Gesamtpunktzahl und Begründungslevel hätte auch in die bestehende Abbildung Abb. 6-2 (S. 124) integriert werden können. Damit verließe man allerdings auch die bivariate Datenanlyse hin zu einem multivariaten Ansatz: Die zweidimensionale Interpretation würde sodann um eine weitere Variable bzw. Achse ergänzt, so dass die im Test erreichte Gesamtpunktzahl, Erklärlevel und Begründungslevel zusammen in einer (noch überschaubaren) dreidimensionalen Darstellung visualisiert werden kann. Die Abbildung Abb. 6-5 (S. 126) darf zum Abschluss kommentarlos den Ausblick in diese Weiten der Datenanalyse geben.

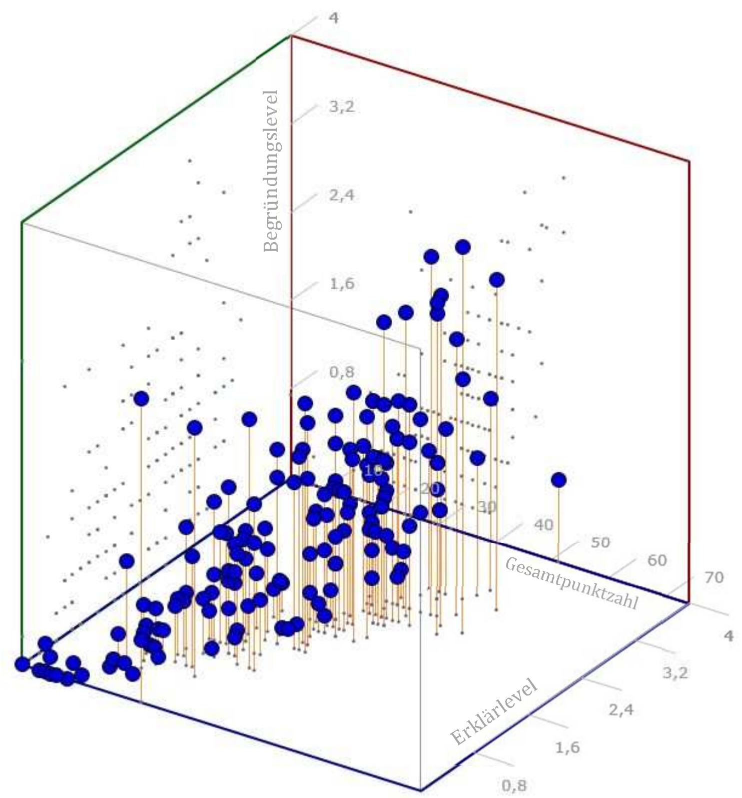


Abb. 6-5: Beispielhafte Darstellung eines dreidimensionalen Modells zur Veranschaulichung einer (möglichen) Korrelation zwischen erreichter Punktzahl, Erklärlevel und Begründungslevel

7 Schlussfolgerungen und Forschungsdesiderata

Wie bereits zu Beginn des Kapitels 2.5 dargestellt, fehlt es nach wie vor an deutschsprachigen Untersuchungen, die die statistische Kompetenz von Schüler/inne/n näher beleuchten. Obwohl die bereits 2003 verabschiedeten Beschlüsse der KMK mittlerweile in allen Bildungs- bzw. Rahmenplänen der einzelnen Bundesländer erfolgreich umgesetzt wurden, fristet das Thema Daten sowohl im Unterricht als auch in der Forschung noch immer ein Schattendasein (vgl. Kap. 2.3). Allein die Verifizierung der in dieser Arbeit nicht untersuchten Ergebnisse und Thesen der im Forschungsstand vorgestellten Studien für den deutschsprachigen Raum könnten für die Mathematikdidaktik und den Unterricht wertvolle Schlussfolgerungen liefern.

Auch eine Öffnung der Einschränkungen, die in diesem Aufgabenset für die vorliegende Untersuchung gemacht wurden, würden viele weitere Forschungsansätze liefern: Wie *gut* erfassen Schüler/innen einen Trend in bivariaten Datensätzen, wenn der Korrelationskoeffizient nicht so deutlich ausgeprägt ist? Entdecken sie eine Korrelation in Datensätzen, wenn diese deutlich größer, die Tabellen unübersichtlicher und nicht nach einem Merkmal vorsortiert angegeben sind? Wie kritisch stehen sie Scheinkorrelationen gegenüber und beziehen den Sachkontext mit ein?

Das zu Beginn dieser Arbeit ausgesprochene Plädoyer, Statistik, insbesondere der verständigen Bearbeitung bivariater Daten, im Unterricht breiteren Raum einzuräumen, sollte eindeutig aufrechterhalten werden. Auch wenn die in dieser Studie untersuchte Prognosefähigkeit bei weitem nicht so schwach ausgebildet ist wie im Forschungsstand angedeutet, haben Jugendliche zumindest beim Thema Daten offenbar deutliche Defizite bzgl. der allgemeinen mathematischen Kompetenzen Begründen (K6) und Kommunizieren (K1). Angesichts der Untersuchungsergebnisse scheint zudem die Frage zulässig, inwieweit Schüler/innen ihre statistischen Fähigkeiten in der Sekundarstufe weiter ausbilden: Sicherlich erweitern sie gegenüber der Primarstufe ihr Faktenwissen bzw. ihr statistisches Repertoire und lernen zumindest, statistische Kenngrößen zu berechnen (vielleicht könnte man dies als *kristalline Statistikintelligenz* bezeichnen). Der Umgang allerdings mit diesen

Kenngrößen, das Bewerten und die Diskussion um deren Aussagekraft, scheint im Unterricht buchstäblich unter den Tisch zu fallen. Didaktiker und Lehrkräfte dürfen angesichts der Ergebnisse dieser Studie deutlich mehr Augenmerk auf die Förderung mathematischen Argumentierens und Begründens legen und dafür im Unterricht weit mehr Zeit anberaumen, als dies aktuell der Fall zu sein scheint. Es bleibt die Frage, wie die in den KMK-Beschlüssen wohl formulierten *ausdifferenzierten* Aspekte dieser allgemeinen mathematischen Kompetenzen künftig im Unterricht gefördert und im Verlauf der Schuljahre an einer weiterführenden Schule aufgebaut werden können.

Interessant wäre übrigens auch, wie erfolgreich Schüler/innen in dem hier verwendeten Test abschneiden würden, wenn zuvor im Unterricht eine entsprechende Unterrichtssequenz zu diesem Thema behandelt worden wäre. Viele Schüler zeigten sich nach den Pretests zu dieser Studie und den anschließenden erklärenden Unterrichtsgesprächen zum einen resigniert („Ach, so wäre das gegangen ...“), aber auch hochinteressiert und motiviert, dieses Themenfeld noch einmal sowohl im Unterricht als auch in einem weiteren Test machen zu dürfen („Können wir das auch mal in der Schule machen? Können wir den Test nochmal schreiben?“).

Diese abschließenden Worte sind für jede Lehrkraft mit einer gewissen Affinität zum Thema Statistik bzw. Daten zwar ermutigend, doch scheinen diese Schülerantworten bedauerlicherweise nach wie vor zu bestätigen, dass das Thema Daten im Unterricht nach wie vor eine eher untergeordnete Rolle spielt. Darum sei zum Schluss noch ein Gedanke aus einem Interview mit D. VOGEL zitiert:

> *„Im Übrigen sind Abstriche an der Statistik nicht im wohlverstandenen Interesse der Schüler, denn sie werden mit statistischen Fragen zumindest als Bürger einer demokratisch verfassten Gesellschaft fortgesetzt konfrontiert werden. Von anderen Inhalten des Mathematikunterrichts kann man dies nicht mit derselben Bestimmtheit sagen.“*

8 Literaturverzeichnis

8.1 Monografien, Zeitschriftenartikel, Aufsätze

BIEHLER R. (2006). Leitidee „Daten und Zufall“ in der didaktischen Konzeption und im Unterrichtsexperiment. MEYER R. (Hrsg.), *Anregungen zum Stochastikunterricht,* 111-120. Hildesheim [u. a.] ; Franzbecker.

BIEHLER, R., HARTUNG, R. (2006). Die Leitidee Daten und Zufall. In BLUM W., DRÜKE-NOE C., R. HARTUNG, KÖLLER O. (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I,* 51-80. Berlin: Cornelsen Scriptor.

BIEHLER R., SCHWEYNOCH S. (1999). Trends und Abweichungen von Trends. *Mathematik lehren, 97*, 17-22.

BIEHLER R., WEBER W. (1995). Entdeckungsreise im Datenland. *Computer + Unterricht, 17,* 4-9.

BÖNIG D., RUWISCH S. (2004). Daten gewinnen, darstellen, verarbeiten und interpretieren. *Die Grundschulzeitschrift, 172,* 6-14.

BOROVCNIK M. (1987). *Materialien zur beschreibenden Statistik und explorativen Datenanalyse.* Wien: Hölder-Pichler-Tempsky.

BURZAN N. (2005). *Quantitative Methoden der Kulturwissenschaften.* Konstanz: UVK Verlagsgesellschaft mbH.

CLEFF T. (2015). *Deskriptive Statistik und Explorative Datenanalyse.* Wiesbaden: Gabler.

CRAMER E., KAMPS U. (2014). *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik.* Berlin: Springer.

EICHLER A. (2006). Spielerlust und Spielerfrust in 50 Jahren Lotto – ein Beispiel für visuell gesteuerte Datenanalyse. *Stochastik in der Schule, 26*, 2-11.

EICHLER A., VOGEL M. (2013). *Leitidee Daten und Zufall.* Wiesbaden: Springer Spektrum.

FRIEL S., BRIGHT G. (1995): Graph Knowledge: Understanding How Students Interpret Data Using Graphs. *Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME) 10/1995.*

FRISCHEMEIER D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots.* Wiesbaden: Springer Spektrum.

GIGENRENZER, G. (1999). *Das Reich des Zufalls. Wissen zwischen Wahrscheinlichkeiten, Häufigkeiten und Unschärfen.* Heidelberg: Spektrum.

GNANADESIKAN R., KETTERING J., SIEGEL A., TUKEY P. (1982). Themen aus der Datenanalyse: Begriffe, Methoden, Beispiele und Pädagogik. *Der Mathematikunterricht 28 (1)*, 28-56.

GONZALEZ T., ESPINEL C., AINLEY J. (2011). Teachers' Graphical Competence. In BATANERO C., BURRILL G., READING C., *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education.* Dordrecht: Springer Netherlands.

GYSIN B. (2013). Kinder erkennen Strukturen. In SPRENGER J., WAGNER A., ZIMMERMANN M. (Hrsg.), *Mathematik lernen, darstellen, deuten, verstehen* (S. 125-138). Wiesbaden, Springer Fachmedien.

HANCOCK, C. (1995). Das Erlernen der Datenanalyse durch anderweitige Beschäftigungen. *Computer und Unterricht, 17,* 33-39. Seelze, Friedrich.

KAUN A. (2006). Stochastik in deutschen Lehrplänen allgemeinbildender Schulen. *Stochastik in der Schule, 26 (3),* 11-17.

KRAUTH J. (1995). Testkonstruktion und Testtheorie. Weinheim: Psychologie VerlagsUnion.

KRÜGER K., SILL H., SIKORA C. (2015). Didaktik der Stochastik in der Sekundarstufe I. Berlin Heidelberg: Springer Spektrum.

MORITZ J. (2004). Reasoning about Covariation. In BEN-ZVI D., GARFIELD J. (Hrsg.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (S. 227-255). Dordrecht: Springer.

NOLL G., SCHMIDT G. (1994). *Interaktive Medien im Unterricht: Trends und Zusammenhänge; Materialien zur Explorativen Datenanalyse und Statistik in der Schule*. Soest: Landesinstitut für Schule und Weiterbildung.

PEREIRA-MENDOZA L. (1995). Graphisches Darstellen von Daten in der Grundschule: Algorithmus kontra Verständnis. *Stochastik in der Schule, 15 (3)*, 5-12. (PEREIRA-MENDOZA L. (1995). Graphing in the Primary School. *Teaching Statistics 17*, 2-6.)

Pereira-Mendoza L., Mellor J. (1990). Students' Concepts of Bar Graphs - Some Preliminary Findings. *ICOTS (International Conference on Teaching Statistic) 3*, 150-157.

Scherrmann, A. (2013): Veranschaulichung statistischer Daten verstehen. In Sprenger J., Wagner A. & Zimmermann M. (Hrsg.), *Mathematik lernen, darstellen, deuten, verstehen* (S. 166-176). Wiesbaden, Springer Fachmedien.

Schwirtz W., Begenat J. (2000). Sind größere Kinder auch schwerer? Ein Statistikprojekt in Klasse 3. *Sache-Wort-Zahl, 28*, 45-50.

Shaughnessy M. (2007). Research on statistics learning and reasoning. In Lester F. K. (Hrsg.), *Second Handbook of Research on Mathematics Teaching and Learning* (S. 957-1010). Charlotte, NC: Information Age Publishing.

Sibbertsen P., Lehne H. (2015). Statistik. Einführung für wirtschafts- und Sozialwissenschaftler. Berlin: Springer.

Stepancik, E. (2012). Trendig! – Linearisierung bivariater Datensätze mit GeoGebra in Klasse 8. *Praxis der Mathematik in der Schule, 44*, 20-23.

Stier, W. (1996). Empirische Forschungsmethoden. Berlin: Springer.

Vogel M., Eichler A. (2010a). Residuen helfen gut zu modellieren. *Stochastik in der Schule, 30 (2)*, 8-13.

Vogel M., Eichler A. (2010b). Leitidee Daten und Zufall in der Sekudarstufe I. In Lindmeier A. (Hrsg.), *Beiträge zum Mathematikunterricht 2010*, 879-882.

Wagner A. (2006). Entwicklung und Förderung von Datenkompetenz in den Klassen 1 - 6. In Biehler R. (Hrsg.), *Kasseler Online-Schriften zur Didaktik der Stochastik* (Band 3). Kassel. Online abrufbar unter http://nbn-resolving.de/urn:nbn:de:hebis:34-2006092214690 (letzter Zugriff am 28.03.2016).

Watson J., Moritz J. (1997). Student analysis of variables in a media context. In Phillips B. (Hrsg.), *Papers on Statistical Education presented at ICMI 8* (S. 129-147). Hawthorn , Australia: Swinburn Press.

Wild C., Pfannkuch M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review, 67 (3)*, 223-248.

Wintermantel G., Vogel D. (2003). *Explorative Datenanalyse – Statistik aktiv lernen*, Stuttgart: Klett.

ZIEFFLER, D. (2008). Learning to Reason About Covariation. In GARFIELD J., BEN-ZVI D. (Hrsg.): *Developing Students' Statistical Reasoning* (S. 289-308). Dordrecht: Springer Netherlands.

8.2 Bildungs- und Rahmenpläne der Länder, KMK, NCTM

BAYERISCHES STAATSMINISTERIUM FÜR BILDUNG UND KULTUS, WISSENSCHAFT UND KUNST (2016). LehrplanPLUS Realschule. München: Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (Hrsg.).

MINISTERIUM FÜR BILDUNG, WISSENSCHAFT UND KULTUR MECKLENBURG-VORPOMMERN (2010), Rahmenplan Mathematik für die Jahrgangsstufen 7 bis 10 des nicht-gymnasialen Bildungsgangs. Schwerin: Ministerium für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern (Hrsg.).

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS (2001). Prinzipien und Standards für Schulmathematik: Datenanalyse und Wahrscheinlichkeit. Übersetzt von Christine Bescherer und Joachim Engel, Ludwigsburg. In M. Borovcnik/ J. Engel/ D. Wickmann (Hrsg.), Anregungen zum Stochastikunterricht: Die NCTM-Standards 2000. Klassische und Bayessche Sichtweise im Vergleich (S. 11-42). Hildesheim, Berlin: Franzbecker Verlag.

NIEDERSÄCHSISCHES KULTUSMINISTERIUM (2014). Kerncurriculum für die Realschule, Schuljahrgänge 5 - 10. Hannover: Niedersächsisches Kultusministerium (Hrsg.).

SEKRETARIAT DER STÄNDIGEN KONFERENZ DER KULTUSMINISTER DER LÄNDER IN DER BUNDESREPUBLIK DEUTSCHLAND (Hrsg.): Bildungsstandards im Fach Mathematik für den Mittleren Bildungsabschluss, Beschluss der Kultusministerkonferenz vom 04.12.2003. München: Wolters Kluwer Deutschland GmbH.

(URL: http://www.brd.nrw.de/lerntreffs/mathe/pages/magazin/leute/vogel.html (Stand: 8.6.2017))

9 Anhang

9.1 Fragebogen

Hauptuntersuchung

TESTHEFT

Daten in Tabellen und Diagrammen

Bitte ausfüllen:

Schule: Realschule ....

Klasse: ☐ 6 ☐ 9 I ☐ 9 II/III

Alter: ....

Geschlecht: ☐ männlich ☐ weiblich

Vorname: ....

Nachname: .... (erster Buchstabe reicht)

A1) Lebenserwartung und Herzfrequenz von Tieren

Wenn in der folgenden Aufgabe z. B. eine *durchschnittliche* Herzfrequenz von 40 angegeben ist, dann bedeutet das, dass das Herz dieses Tieres normalerweise 40× pro Minute schlägt.
(Natürlich kann das Herz auch mal etwas schneller oder langsamer schlagen, denn auch ein Tier kann mal aufgeregt sein oder muss ein wenig schlafen.)

Ebenso verhält es sich mit der *durchschnittlichen* Lebenserwartung: Ein Löwe z. B. hat nicht immer eine exakte Lebenserwartung von 23 Jahren. Manche Löwen werden ein bisschen älter, manche versterben ein wenig früher. Aber im Schnitt werden sie 23 Jahre alt.

	Tier	**durchschnittliche Anzahl der Herzschläge** (pro Minute)	**durchschnittliche Lebenserwartung** (in Jahren)
1	Löwe	40	23
2	See-Elefant (an der Wasseroberfläche)	60	20
3	Ziege	68	18
4	Schaf	70	16
5	Storch	85	18
6	Fuchs	100	13
7	Hauskatze	120	12
8	Kaninchen	215	9
9	Meerschweinchen	256	10
10	Huhn	353	6
11	Krähe	380	3
12	Goldhamster	425	2

☞ Schau dir nun in Ruhe die obige Tabelle an und beantworte <u>mit Hilfe dieser Tabelle</u> folgende Fragen:

1 a) Das Herz eines Kaninchens schlägt durchschnittlich 215 Mal pro Minute.
☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

1 b) Schreibe in das Kästchen, welche Lebenserwartung ein Fuchs durchschnittlich hat:

1 c) Ein Storch hat eine durchschnittlich höhere Lebenserwartung als ein Huhn.

☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

1 d) Die durchschnittliche Herzfrequenz eines Meerschweinchens ist höher als die eines Goldhamsters.

☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

1 e) Eine Giraffe hat eine durchschnittliche Herzfrequenz von 75.
Was vermutest du, wie alt so eine Giraffe wird?

Antwort: ☐

Erkläre möglichst genau, wie du auf diese Zahl kommst.

....

....

....

....

....

....

....

→ *Bitte umblättern!*

2) Gehalt und Krankentage

Die folgende Tabelle zeigt das durchschnittliche Monatseinkommen von Personen verschiedener Berufsgruppen sowie die Anzahl der Tage, an denen Personen dieser Berufsgruppe durchschnittlich im Jahr krank gemeldet waren.

	Branche	**durchschnittliches Einkommen pro Monat** (in €)	**durchschnittliche Krankentage pro Jahr** (gerundet)
1	Gastgewerbe	1300	11
2	Baugewerbe	1800	9
4	Verkehr und Logistik	2000	8
3	Handel	2200	6
5	Gesundheitswesen	2500	7
6	Öffentliche Verwaltung	2700	8
7	Produzierendes Gewerbe	2900	5
8	Erziehung und Unterricht	3000	4
9	freiberufl., techn. oder wiss. Dienstleistungen	3500	4
10	Energieversorgung	4000	1
11	Information und Kommunikation	4700	2
12	Finanzen und Versicherung	5700	1

☞ Schau dir erst wieder die obige Tabelle an und beantworte dann mit Hilfe der Tabelle folgende Fragen:

2 a) Personen, die im Gesundheitswesen angestellt sind, verdienen durchschnittlich 2700 € pro Monat.

☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

2 b) Schreibe in das Kästchen, wie viele Krankheitstage eine Person aus dem Bereich „Information und Kommunikation" durchschnittlich pro Jahr hat:

2 c) Personen aus dem Bereich „Erziehung und Unterricht" verdienen durchschnittlich weniger als Personen, die im Gesundheitswesen beschäftigt sind.

☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

2 d) Personen aus dem Baugewerbe sind durchschnittlich öfter krank gemeldet als Personen aus dem Gastgewerbe?

☐ richtig ☐ falsch ☐ Das kann man mit Hilfe dieser Tabelle nicht bestimmen.

2 e) Der Bereich „Grundstücks- und Wohnungswesen" ist nicht in der Tabelle aufgeführt. Eine Person aus diesem Bereich verdient durchschnittlich 2600 € pro Monat.

Was vermutest du, wie viele Krankentagen eine Person aus dieser Berufsgruppe hat?

Antwort: ☐

Erkläre, wie du auf diese Anzahl kommst.

....

....

....

....

....

2 f) Kreuze an, welche Aussagen du mit Hilfe des Diagramms machen kannst:

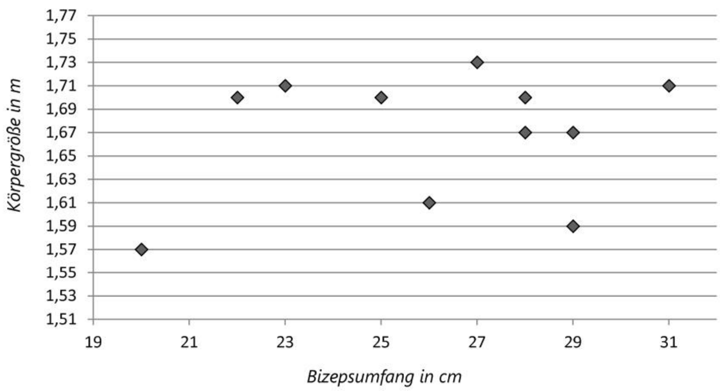
☐ Je mehr man durchschnittlich verdient, desto häufiger ist man krank.
☐ Wer mehr verdient, ist seltener krank.
☐ Je mehr Krankentage man durchschnittlich hat, desto weniger verdient man.
☐ Zwischen Einkommen und Krankentagen besteht kein Zusammenhang.

→ *Bitte umblättern!*

kK-G-☒

3) Dicker Bizeps

Bei elf Schülern der Klasse 8 d (nur Jungen) wurde sowohl die Körpergröße als auch der Umfang des Oberarms gemessen.



☞ Schau dir das Diagramm in Ruhe an und überlege:

3 a) Andreas aus der Parallelklasse (8 b) ist 1,67 m groß. Welchen Umfang könnte sein Oberarm haben?

- ☐ Der müsste ungefähr 24 cm haben.
- ☐ 28 oder 29 cm.
- ☐ Auf alle Fälle mindestens 28 cm!
- ☐ Das lässt sich mit Hilfe des Diagramms nicht beantworten. Der Oberarm könnte auch dicker oder dünner sein als im Diagramm dargestellt.

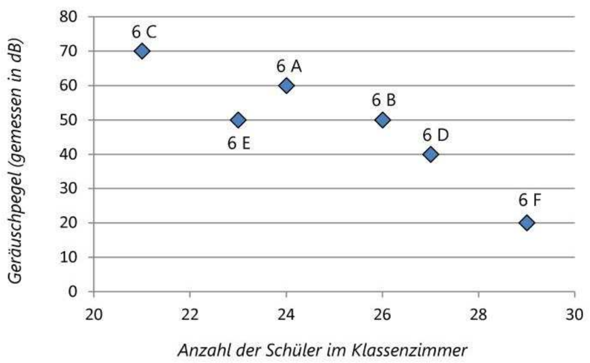
3 b) Marina behauptet: „Je größer ein Junge ist, desto dicker ist sein Oberarm!“
Was sagst du dazu? Kreuze an.

- ☐ Ja, das stimmt. Das Diagramm bestätigt Marinas Vermutung.
- ☐ Nein, das Gegenteil ist der Fall: Je größer man ist, desto dünner sind normalerweise die Oberarme.
- ☐ Aus dem Diagramm lässt sich keine „Gesetzmäßigkeit“ darüber ableiten.

4) Lärm im Klassenzimmer

nK-G-...

Eine kleine Gruppe von Sechstklässlern führt ein Projekt zum Thema „Geräusche und Lärm in der Schule" durch. Dazu messen sie auch den Geräuschpegel in sechs verschiedenen Klassenzimmern. Sie benutzen dazu ein so genanntes Schallpegelmessgerät. In jedem Klassenzimmer schreiben sie den Wert auf, den das Gerät anzeigt, und zählen die Anzahl der Schüler, die gerade in diesem Klassenzimmer Unterricht haben. Ihr Ergebnis stellt die kleine Forschungsgruppe in folgendem Diagramm dar:



☞ Beantworte <u>mit Hilfe des Diagramms</u> die folgenden beiden Fragen:

4 a) Angenommen, die kleine Forschungsgruppe führt eine weitere Schallpegelmessung in einer Klasse mit 22 Schülern durch. Welchen Wert würden sie deiner Meinung nach erhalten?

Antwort: ☐

Erkläre, wie du auf diesen Wert kommst.

....

....

....

4 b) Max behauptet: „Das Diagramm zeigt, dass in einem Klassenzimmer mit mehr Personen weniger Lärm herrscht." Was meinst du dazu? Begründe deine Antwort so ausführlich wie möglich.

....

....

....

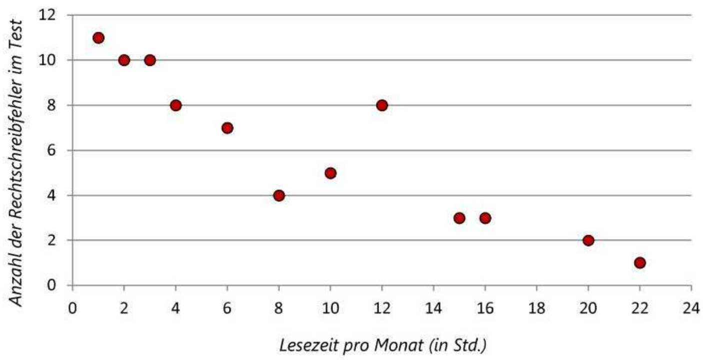
....

....

nK-T-☒

5. Lesen und Rechtschreibung

Zwölf Schüler wurden *vor* einem Rechtschreibtest befragt, wie viel Stunden sie ungefähr pro Monat lesen. Anschließend wurde der Test ausgewertet und das Ergebnis grafisch dargestellt.



☞ Schau dir das Diagramm in Ruhe an und überlege:

5 a) Ein anderer Schüler, der bei diesem Test nicht mitgeschrieben hat (er war an diesem Tag leider krank), gibt an, dass er ca. 11 Stunden pro Monat liest. Was glaubst du, wie viele Rechtschreibfehler er wohl haben würde, wenn er auch diesen Test schreibt?

Antwort: ☐

Erkläre, wie du auf diese Anzahl kommst.

....

....

....

....

5 b) Kreuze an, welche Aussagen du mit Hilfe des Diagramms machen kannst:

- ☐ Wer mehr liest, macht auch durchschnittlich mehr Fehler.
- ☐ Je mehr Stunden ein Schüler durchschnittlich liest, desto weniger Fehler macht er im Test.
- ☐ Je größer die Fehleranzahl im Rechtschreibtest, desto geringer ist die durchschnittliche Lesezeit pro Monat.
- ☐ Zwischen der Leistung im Rechtschreibtest und der Anzahl der Stunden, die ein Schüler pro Monat liest, besteht kein Zusammenhang.

6) Wer ist der Schnellste?

In einem Reaktionstest wird auf einem Bildschirm eine Farbe angezeigt. Die Testperson soll dann so schnell wie möglich den Knopf mit der richtigen Farbe drücken. Erst wenn der richtige Knopf gedrückt wurde, wird die nächste Farbe angezeigt.

Die nebenstehende Tabelle zeigt, wie lange Testpersonen verschiedenen Alters durchschnittlich gebraucht haben, um den richtigen Knopf zu finden.

	Reaktionszeit in Sekunden	**Alter** in Jahren
Testperson 1	0,40	12
Testperson 2	0,45	16
Testperson 3	0,50	16
Testperson 4	0,55	17
Testperson 5	0,60	20
Testperson 6	0,65	26
Testperson 7	0,65	38
Testperson 8	0,75	32
Testperson 9	0,80	40
Testperson 10	0,90	45
Testperson 11	1,00	35
Testperson 12	1,30	60
Testperson 13	1,30	72
Testperson 14	1,35	55

☞ Schau dir die Tabelle an und überlege:

6 a) Eine weitere Testperson im Alter von 42 Jahren macht denselben Reaktionstest. Was schätzt du, wie viele Sekunden diese Person durchschnittlich braucht um den richtigen Knopf zu drücken?

Antwort: ☐

Erkläre, wie du auf diesen Wert kommst.

....

....

....

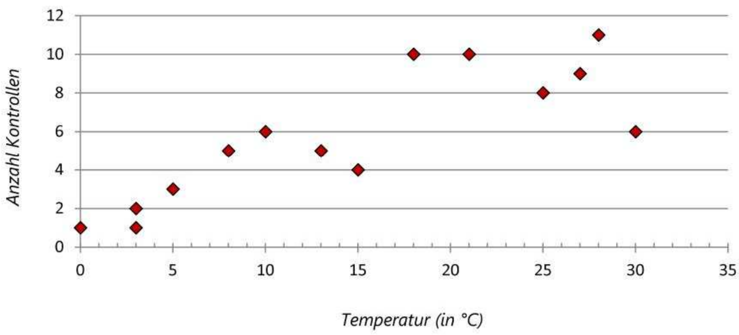
....

6 b) Kreuze an, welche Aussagen du mit Hilfe der Tabelle machen kannst:

- ☐ Je älter die Testperson, desto kürzer ist die Reaktionszeit.
- ☐ Ältere Testpersonen brauchen in der Regel länger als jüngere Testteilnehmer.
- ☐ Grob gesagt gilt die Regel: Je länger man braucht, desto älter ist man.
- ☐ Aus der Tabelle kann man keinen Zusammenhang bzw. keine Regel zwischen dem Alter und der Reaktionszeit rauslesen. Die Tabelle zeigt: Auch im Alter kann man reaktionsschnell sein und jüngere Menschen können auch mal langsam sein.

7) Frühlingshafte Radarkontrollen

Herr Müller wundert sich, dass Geschwindigkeitskontrollen im Winter seltener stattfinden als im Sommer. Er hat auf seinem Handy eine App installiert, die ihm die Anzahl der Geschwindigkeitskontrollen in seiner Heimatstadt anzeigt. Über einen längeren Zeitraum hat er die Anzahl der Kontrollen und die Mittagstemperatur notiert. Seine Ergebnisse hat er in folgendem Diagramm dargestellt:



☞ Schau dir das Diagramm an und überlege:

7 a) An einem sonnigen Tag hat es mittags 14° C. Was denkst du, wie viele Verkehrskontrollen auf Herrn Müllers Handy-App angezeigt werden?

Antwort: [ ]

Erkläre, wie du auf diesen Wert kommst.

....

....

....

....

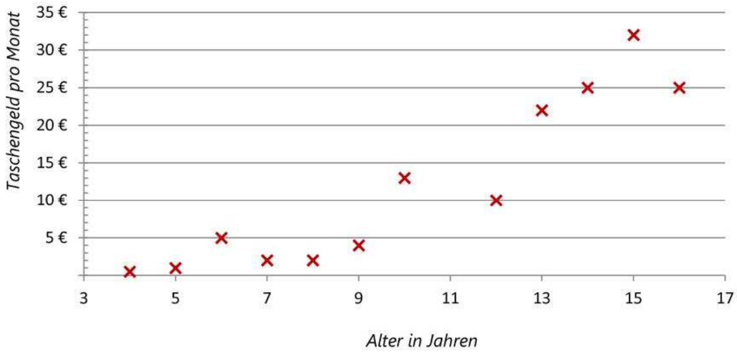
7 b) Kreuze jeweils an, ob du die Aussage mithilfe des Diagramms machen kannst:

- ☐ Je mehr Blitzer man am Straßenrand sieht, desto wärmer ist es draußen.
- ☐ Je wärmer es draußen ist, desto weniger Kontrollen macht die Polizei.
- ☐ Je wärmer es ist, desto häufiger wird kontrolliert.
- ☐ In der Grafik kann man nicht erkennen, ob die Anzahl der Polizeikontrollen mit der Außentemperatur zusammenhängt.

pK-G-...

8) Mehr Taschengeld!

Das Deutsche Jugendinstitut gibt jedes Jahr aktualisierte Empfehlungen heraus, wie viel Taschengeld pro Woche für Kinder und Jugendliche angemessen sind. Ob sich die Eltern daran halten, bleibt natürlich ihnen überlassen. Eine schnelle Umfrage mit ein paar Schülern an der Bushaltestelle ergab folgendes Bild:



☞ Schau dir das Diagramm an und überlege:

8 a) Offenbar wurde kein Elfjähriger zu diesem Thema befragt. Was würdest du vermuten, wie viel Taschengeld er pro Woche bekommt?

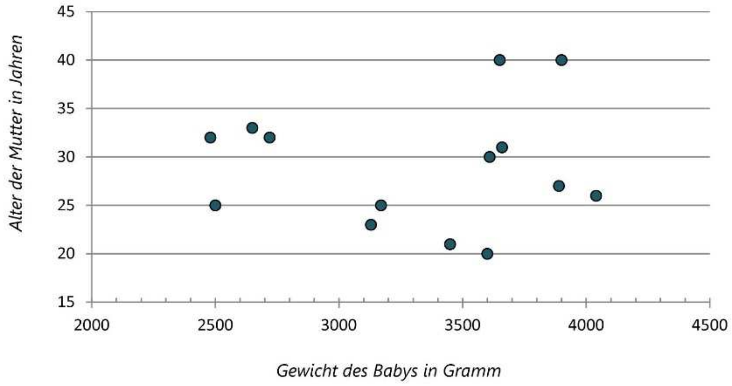
Antwort:

Erkläre, wie du auf diesen Betrag kommst.

8 b) Jana behauptet: „Ist doch eh klar! Je älter man ist, desto mehr Taschengeld bekommt man!" Was meinst du dazu? Begründe deine Antwort mit Hilfe des Diagramms.

9) Babys und Mamas

Bei der Geburt eines Babys werden viele Daten über das Baby und seine Mutter erhoben. Im folgenden Diagramm zum Beispiel ist das Alter von Müttern sowie das Gewicht ihres frisch geborenen Babys dargestellt.



☞ Schau dir das Diagramm an und überlege:

9 a) Ein weiteres Baby wiegt bei seiner Geburt genau 3400 Gramm. Wie alt könnte die Mutter dieses Babys sein?

- ☐ Die müsste ca. 30 Jahre alt sein.
- ☐ Zwischen 21 und 23.
- ☐ Auf alle Fälle älter als 20 und jünger als 40.
- ☐ Diese Frage kann man mit Hilfe des Diagramms nicht beantworten. Die Mutter könnte auch jünger oder älter sein als im Diagramm dargestellt.

9 b) Erkennst du im obigen Diagramm einen Zusammenhang zwischen dem Alter der Mutter und dem Gewicht des Babys? Begründe deine Antwort.

....

....

....

....

10) Tiefkühlkost und Stromverbrauch

Hast du gewusst, dass jedes Jahr am 6. März der Tag der Tiefkühlkost gefeiert wird?
Seit gut 60 Jahren schon werden nun tiefgekühlte Lebensmittel verkauft. Damit diese zuhause weiter aufbewahrt werden können, braucht man natürlich eine Gefriertruhe oder einen Gefrierschrank. Und die brauchen Strom ...

	Jahr	**jährlicher Konsum von Tiefkühlkost pro Kopf** (in kg; gerundet)	**durchschnittl. jährlicher Stromverbrauch eines Haushalts** (in kWh)
1	1998	29,0	2980
2	1999	31,5	2990
3	2000	33,0	3102
4	2001	34,0	3100
5	2003	35,0	3090
6	2004	36,0	3220
7	2005	37,0	3270
8	2006	38,0	3320
9	2008	39,0	3310
10	2009	39,5	3210
11	2010	41,0	3290
12	2012	40,0	3250

☞ Schau dir die Tabelle genau an und beantworte dann die beiden untenstehenden Fragen.

10 a) Im Jahr 2002 verspeiste der deutsche Durchschnittsbürger 34,5 kg Tiefkühlkost. Welchen Stromverbrauch vermutest du in diesem Jahr in einem deutschen Durchschnittshaushalt?

Antwort:

Erkläre, wie du auf diesen Wert kommst.

....

....

....

....

10 b) Erkennst du in der obigen Tabelle einen Zusammenhang zwischen der verkauften Menge an Tiefkühlkost und dem Stromverbrauch in deutschen Haushalten? Begründe deine Antwort.

....

....

....

....

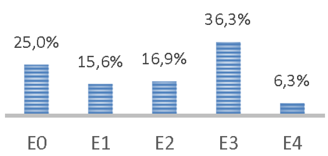
9.2 Deskriptive Auswertung der Fragebögen

Aufgabe 1 (Lebenserwartung und Herzfrequenz von Tieren)

Korrelationstyp: negativ
Darstellungsform: Tabelle

		korrekte Antworten (N = 160):	
1 a)	*Das Herz eines Kaninchens schlägt durchschnittlich 215 Mal pro Minute.* **(richtig)**	158	(98,8 %)
1 b)	*Schreibe in das Kästchen, welche Lebenserwartung ein Fuchs durchschnittlich hat.* **(13)**	158	(98,8 %)
1 c)	*Ein Storch hat eine durchschnittlich höhere Lebenserwartung als ein Huhn.* **(richtig)**	153	(95,6 %)
1 d)	*Die durchschnittliche Herzfrequenz eines Meerschweinchens ist höher als die eines Goldhamsters.* **(falsch)**	137	(85,6 %)
1 e)	*Eine Giraffe hat eine durchschnittliche Herzfrequenz von 75. Was vermutest du, wie alt so eine Giraffe wird?* **(korrekt: Wert zw. [13; 20] + Erklärung)**		

Häufigste Angabe: $x_{Mod} = 17$ (87×)
Mittelwert: $\bar{x} = 18{,}4$



Häufigkeit der Erklärlevel:

E0:	40	(25,0 %)
E1:	25	(15,6 %)
E2:	27	(16,9 %)
E3:	58	(36,3 %)
E4:	10	(6,3 %)

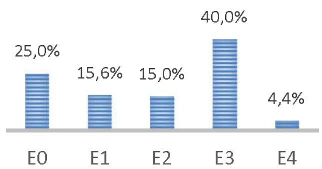
→

Aufgabe 2 (Gehalt und Krankentage)

Korrelationstyp: negativ
Darstellungsform: Tabelle

		korrekte Antworten (N = 160):	
2 a)	*Personen, die im Gesundheitswesen angestellt sind, verdienen durchschnittlich 2700 € pro Monat.* **(falsch)**	145	(90,6 %)
2 b)	*Schreibe in das Kästchen, wie viele Krankheitstage eine Person aus dem Bereich „Information und Kommunikation" durchschnittlich pro Jahr hat.* **(2)**	155	(96,9 %)
2 c)	*Personen aus dem Bereich „Erziehung und Unterricht" verdienen durchschnittlich weniger als Personen, die im Gesundheitswesen beschäftigt sind.* **(falsch)**	139	(86,9 %)
2 d)	*Personen aus dem Baugewerbe sind durchschnittlich öfter krank gemeldet als Personen aus dem Gastgewerbe?* **(falsch)**	147	(91,9 %)
2 e)	*Der Bereich „Grundstücks- und Wohnungswesen" ist nicht in der Tabelle aufgeführt. Eine Person aus diesem Bereich verdient durchschnittlich 2600 € pro Monat. Was vermutest du, wie viele Krankentage eine Person aus dieser Berufsgruppe hat?* **(korrekt: Wert zw. [5; 8])**		

Häufigste Angabe: $x_{Mod} = 7{,}5$ (51×)
Mittelwert: $\bar{x} = 7{,}2$



Häufigkeit der Erklärlevel:

E0:	40	(25,0 %)
E1:	25	(15,6 %)
E2:	24	(15,0 %)
E3:	64	(40,0 %)
E4:	7	(4,4 %)

2 f) *Kreuze an, welche Aussagen du mit Hilfe des Diagramms machen kannst.*

(richtig: ☒ **Wer mehr verdient, ist seltener krank.**

☒ **Je mehr Krankentage man durchschnittlich hat, desto weniger verdient man.)**

Vollständig korrekt gelöst (≙ 2 P):	79	(49,4 %)
1 Kreuzchen richtig (aber keine widersprüchlichen Antworten, ≙ 1 P):	29	(18,1 %)

→

Aufgabe 3 (Dicker Bizeps)

Korrelationstyp: ohne

Darstellungsform: Diagramm

3 a) *Andreas aus der Parallelklasse (8 b) ist 1,67 m groß. Welchen Umfang könnte sein Oberarm haben?*

(richtig: ☒ Das lässt sich mit Hilfe des Diagramms nicht beantworten. Der Oberarm könnte auch dicker oder dünner sein als im Diagramm dargestellt.)

Korrekt gelöst (≙ 1 P): 82 (51,3 %)

3 b) *Marina behauptet: „Je größer ein Junge ist, desto dicker ist sein Oberarm!“*

Was sagst du dazu? Kreuze an.

(richtig: ☒ Aus dem Diagramm lässt sich keine „Gesetzmäßigkeit“ darüber ableiten.)

Korrekt gelöst (≙ 1 P): 134 (83,8 %)

Aufgabe 4 (Lärm im Klassenzimmer)

Korrelationstyp: negativ

Darstellungsform: Diagramm

Häufigkeit der „Erklärlevel“ (N = 160):

4 a) *Angenommen, die kleine Forschungsgruppe führt eine weitere Schallpegelmessung in einer Klasse mit 22 Schülern durch. Welchen Wert würden sie deiner Meinung nach erhalten?* **(korrekt: Wert zw. [50; 70])**

E0:	62	(38,8 %)
E1:	28	(17, 5%)
E2:	26	(16,3 %)
E3:	33	(20,6 %)
E4:	11	(6,9 %)

Häufigste Angabe: $x_{Mod} = 60$ (41×)

Mittelwert: $\bar{x} = 50{,}4$

Häufigkeit der „Begründungslevel“ (N = 160):

4 b) *Max behauptet: „Das Diagramm zeigt, dass in einem Klassenzimmer mit mehr Personen weniger Lärm herrscht.“ Was meinst du dazu? Begründe deine Antwort so ausführlich wie möglich.*

B0:	89	(55,6 %)
B1:	25	(15,6 %)
B2:	15	(9,4 %)
B3:	14	(8,8 %)
B4:	17	(10,6 %)

→

Aufgabe 5 (Lesen und Rechtschreibung)

Korrelationstyp: negativ

Darstellungsform: Diagramm

Häufigkeit der „Erklärlevel“ (N = 160):

5 a) *Ein anderer Schüler, der bei diesem Test nicht mitgeschrieben hat, gibt an, dass er ca. 11 Stunden pro Monat liest. Was glaubst du, wie viele Rechtschreibfehler er wohl haben würde, wenn er auch diesen Test schreibt?*
(korrekt: Wert zw. [3; 7])

E0:	43	(26,9 %)
E1:	30	(18,8 %)
E2:	27	(16,9 %)
E3:	42	(26,3 %)
E4:	18	(11,3 %)

Häufigste Angabe: $x_{Mod} = 6$ (32×)
Mittelwert: $\bar{x} = 5{,}5$

5 b) *Kreuze an, welche Aussagen du mit Hilfe des Diagramms machen kannst.*

(richtig:
☒ **Je mehr Stunden ein Schüler durchschnittlich liest, desto weniger Fehler macht er im Test.**
☒ **Je größer die Fehleranzahl im Rechtschreibtest, desto geringer ist die durchschnittliche Lesezeit pro Monat.)**

Vollständig korrekt gelöst (≙ 2 P):	79	(49,4 %)
1 Kreuzchen richtig (aber keine widersprüchlichen Antworten, ≙ 1 P):	28	(17,5 %)

Aufgabe 6 (Wer ist der Schnellste?)

Korrelationstyp: positiv

Darstellungsform: Tabelle

Häufigkeit der „Erklärlevel“ (N = 160):

6 a) *Eine weitere Testperson im Alter von 42 Jahren macht denselben Reaktionstest. Was schätzt du, wie viele Sekunden diese Person durchschnittlich braucht um den richtigen Knopf zu drücken?* **(korrekt: Wert zw. [0,8; 1,1])**

E0:	29	(18,1 %)
E1:	21	(13,1 %)
E2:	16	(10,0 %)
E3:	89	(55,6 %)
E4:	5	(3,1 %)

Häufigste Angabe: $x_{Mod} = 0{,}85$ (86×)
Mittelwert: $\bar{x} = 0{,}93$

Häufigkeit der „Begründungsstufen“ (N = 160):

6 b) *Kreuze an, welche Aussagen du mit Hilfe des Diagramms machen kannst.*

(richtig:
☒ **Ältere Testpersonen brauchen in der Regel länger als jüngere Testteilnehmer.**
☒ **Grob gesagt gilt die Regel: Je länger man braucht, desto älter ist man.)**

Vollständig korrekt gelöst (≙ 2 P):	72	(45,0 %)
1 Kreuzchen richtig (aber keine widersprüchlichen Antworten, ≙ 1 P):	32	(20,0 %)

Aufgabe 7 (Frühlingshafte Radarkontrollen)

Korrelationstyp: positiv

Darstellungsform: Diagramm

Häufigkeit der „Erklärlevel" (N = 160):

7 a) *An einem sonnigen Tag hat es mittags 14° C. Was denkst du, wie viele Verkehrskontrollen auf Herrn Müllers Handy-App angezeigt werden?*
(korrekt: Wert zw. [4; 8])

E0:	38	(23,8 %)
E1:	38	(23,8 %)
E2:	28	(17,5 %)
E3:	45	(29,4 %)
E4:	9	(5,6 %)

Häufigste Angabe: $x_{Mod} = 5$ (55×)
Mittelwert: $\bar{x} = 5{,}3$

7 b) *Kreuze jeweils an, ob du die Aussage mit Hilfe des Diagramms machen kannst.*
(richtig: ☒ Je mehr Blitzer man am Straßenrand sieht, desto wärmer ist es draußen.
☒ Je wärmer es ist, desto häufiger wird kontrolliert.)

Vollständig korrekt gelöst (≙ 2 P):	49	(30,6 %)
1 Kreuzchen richtig (aber keine widersprüchlichen Antworten, ≙ 1 P):	40	(25,0 %)

Aufgabe 8 (Mehr Taschengeld!)

Korrelationstyp: positiv

Darstellungsform: Diagramm

Häufigkeit der „Erklärlevel" (N = 160):

8 a) *Offenbar wurde kein Elfjähriger zu diesem Thema befragt. Was würdest du vermuten, wie viel Taschengeld er pro Woche bekommt?*
(korrekt: Wert zw. [5; 20])

E0:	45	(28,1 %)
E1:	41	(25,6 %)
E2:	21	(13,1 %)
E3:	49	(30,6 %)
E4:	4	(2,5 %)

Häufigste Angabe: $x_{Mod} = 12$ (31×)
Mittelwert: $\bar{x} = 11{,}4$

Häufigkeit der „Begründungslevel" (N = 160):

8 b) *Jana behauptet: „Ist doch eh klar! Je älter man ist, desto mehr Taschengeld bekommt man!" Was meinst du dazu? Begründe deine Antwort mit Hilfe des Diagramms.*

B0:	68	(42,5 %)
B1:	52	(32,5 %)
B2:	8	(5,0 %)
B3:	19	(11,9 %)
B4:	13	(8,1 %)

→

Aufgabe 9 (Babys und Mamas)

Korrelationstyp: ohne

Darstellungsform: Diagramm

9 a) *Ein weiteres Baby wiegt bei seiner Geburt genau 3400 Gramm. Wie alt könnte die Mutter dieses Babys sein?*

(richtig: ☒ Diese Frage kann man mit Hilfe des Diagramms nicht beantworten. Die Mutter könnte auch jünger oder älter sein als im Diagramm dargestellt.)

Korrekt gelöst (≙ 1 P): 64 (40,0 %)

Häufigkeit der „Begründungslevel" (N = 160):

9 b) *Erkennst du im obigen Diagramm einen Zusammenhang zwischen dem Alter der Mutter und dem Gewicht des Babys? Begründe deine Antwort.*

B0:	61	(38,1 %)
B1:	82	(51,3 %)
B2:	3	(1,9 %)
B3:	6	(3,8 %)
B4:	8	(5,0 %)

Aufgabe 10 (Tiefkühlkost und Stromverbrauch)

Korrelationstyp: positiv

Darstellungsform: Tabelle

Häufigkeit der „Erklärlevel" (N = 160):

10 a) *Im Jahr 2002 verspeiste der deutsche Durchschnittsbürger 34,5 kg Tiefkühlkost. Welchen Stromverbrauch vermutest du in diesem Jahr in einem deutschen Durchschnittshaushalt?* **(korrekt: Wert zw. [3090; 3200])**

E0:	59	(36,9 %)
E1:	25	(15,6 %)
E2:	10	(6,3 %)
E3:	65	(40,6 %)
E4:	1	(0,6 %)

Häufigste Angabe: $x_{Mod} = 3095$ (57×)
Mittelwert: $\bar{x} = 3067{,}4$

Häufigkeit der „Begründungslevel" (N = 160):

10 b) *Erkennst du in der obigen Tabelle einen Zusammenhang zwischen der verkauften Menge an Tiefkühlkost und dem Stromverbrauch in deutschen Haushalten? Begründe deine Antwort.*

B0:	110	(68,8 %)
B1:	14	(8,8 %)
B2:	0	(0,0 %)
B3:	6	(3,8 %)
B4:	30	(18,8 %)

9.3 CD

Sie enthält neben der Arbeit den Fragebogen sowie die umfangreichen Excel-Tabellen zur Auswertung.