

## Statistik in der Schule



## Zu den Autorinnen



### Dr. Michaela Schmidt:

Geboren am 3. August 1976 in Eggenfelden, 1991 bis 1993 Ausbildung zur Kinderpflegerin, 1993 bis 1997 berufliche Tätigkeit als Kinderpflegerin, 1997 bis 2000 Berufsoberschule, 2000 bis 2005 Studium der Psychologie an der Technischen Universität Darmstadt mit den Schwerpunkten Pädagogische Psychologie und Arbeits- und Organisationspsychologie, 2009 Promotion, 2005 bis 2009 Wissenschaftliche Mitarbeiterin an der Technischen Universität Darmstadt im Bereich der Pädagogischen Psychologie.

Seit 2009 Wissenschaftliche Mitarbeiterin am Lehrstuhl für Empirische Schul- und Unterrichtsforschung an der Universität des Saarlandes in Saarbrücken.



### Prof. Dr. Franziska Perels:

Geboren am 13. August 1973 in Bad Karlshafen. 1994 bis 2000 Studium der Psychologie an der Technischen Universität Berlin mit den Schwerpunkten Pädagogische Psychologie und Kognitionspsychologie, 2002 Promotion an der Technischen Universität Darmstadt, 2000 bis 2005 Wissenschaftliche Mitarbeiterin an der Technischen Universität Darmstadt im Bereich Pädagogische Psychologie, 2005 bis 2009 Arbeitseinheitsleiterin und stellvertretende Abteilungsleiterin am Institut für Qualitätsentwicklung in Wiesbaden, 2009 Habilitation im Fach Psychologie.

Seit 2009 Professorin für Empirische Schul- und Unterrichtsforschung an der Universität des Saarlandes in Saarbrücken.

---

## Impressum

- Herausgeber:** Institut für Qualitätsentwicklung (IQ),  
Walter-Hallstein-Straße 5-7, 65197 Wiesbaden  
Tel.: (0611) 5827-0  
Fax: (0611) 5827-109  
E-Mail: [info@iq.hessen.de](mailto:info@iq.hessen.de)  
Internet: [www.iq.hessen.de](http://www.iq.hessen.de)
- Autorinnen:** Prof. Dr. Franziska Perels  
Dr. Michaela Schmidt
- Redaktion:** Dr. Reinhold Fischenich
- Gestaltung:** [s.tietze@medien-frankfurt.com](mailto:s.tietze@medien-frankfurt.com)
- Fotos:** Ewa Walicka © [www.fotolia.de](http://www.fotolia.de), Dmitry Sunagatov © [www.fotolia.de](http://www.fotolia.de), herreneck © [www.fotolia.de](http://www.fotolia.de), Yang MingQi © [www.fotolia.de](http://www.fotolia.de), womue © [www.fotolia.de](http://www.fotolia.de), Oleksandr © [www.fotolia.de](http://www.fotolia.de), Dr. Reinhold Fischenich
- Druck:** Druckerei der Justizvollzugsanstalt Darmstadt
- 1. Auflage:** September 2009
- Hinweis:** Als Online-Fassung finden Sie diesen Text auch auf der Internetseite des Instituts für Qualitätsentwicklung unter [www.iq.hessen.de](http://www.iq.hessen.de)

Diese Druckschrift wird im Rahmen der Öffentlichkeitsarbeit der Hessischen Landesregierung herausgegeben. Sie stellt jedoch keine verbindliche, amtliche Verlautbarung des Hessischen Kultusministeriums dar. Vielmehr will sie zur Diskussion über die behandelten Themen anregen und zur Weiterentwicklung des hessischen Schulwesens beitragen. Dem Land Hessen (Institut für Qualitätsentwicklung) sind an den abgedruckten Beiträgen alle Rechte an der Veröffentlichung, Verbreitung, Übersetzung und auch die Einspeicherung und Ausgabe in Datenbanken vorbehalten.

# Inhalt

---

<b>1. Einleitung: Statistik in der Schule</b>	4
<b>2. Praxisbeispiele und deren Interpretation</b>	6
2.1 Ergebnisse interner und externer Evaluation von Schule	6
2.2 Zentrale Lernstandserhebungen	9
2.3 (Internationale) Schulleistungsstudien	11
<b>3. Schulrelevante statistische Kennwerte und Verfahren</b>	13
3.1 Skalenniveaus	13
3.2 Statistische Kennwerte	15
3.3 Darstellungsformen	19
3.4 Statistische Bedeutsamkeit: Signifikanz	21
3.5 Zusammenhangsmaße	23
3.6 Grenzen und Probleme	24
<b>4. Fazit</b>	27

# 1. Einleitung: Statistik in der Schule

## Warum haben Sie diese Broschüre in die Hand genommen?

Es gibt ja verschiedene Gründe, zu einem solchen Heft zu greifen: Sind Sie inhaltlich an dem Thema „Statistik und Evaluation“ interessiert? Oder benötigen Sie eher eine Lesehilfe für die statistischen Daten, mit denen Sie in Ihrem Beruf zurzeit zunehmend stärker konfrontiert werden? Oder sind Sie einigermaßen bewandert in der Auswertung und Interpretation statistischer Daten und möchten zur Sicherheit noch einmal bestimmte Begrifflichkeiten nachschlagen?

Die vorliegende Broschüre möchte all diesen Anliegen gerecht werden und verfolgt drei Ziele:

### Ziele dieser Broschüre

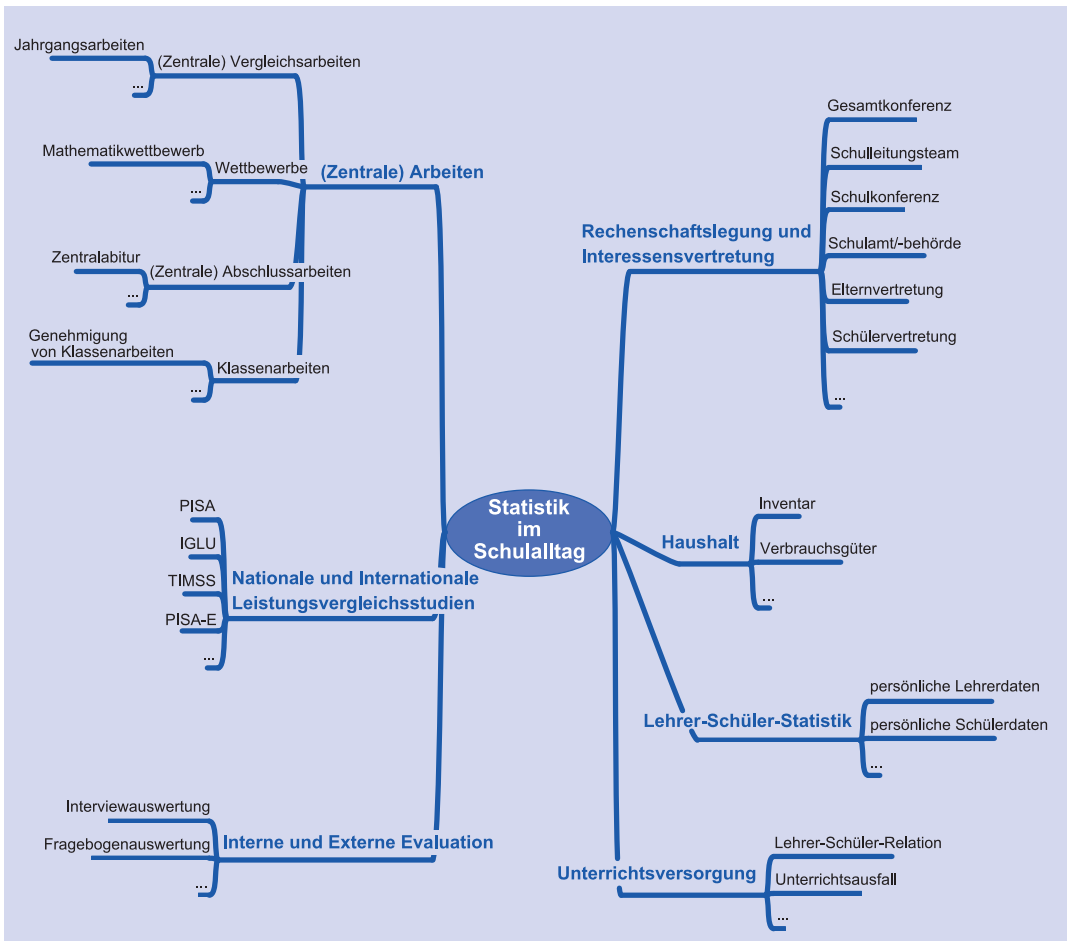
- Zunächst werden beispielhaft drei Bereiche des Schulwesens umrissen, für die es notwendig ist, über eine Grundvorstellung von Statistik und Evaluation zu verfügen.
- Anschließend wird ein detaillierter Überblick über statistische Kennwerte und Verfahren gegeben. Dabei werden relativ unabhängig vom schulischen Kontext statistische Grundbegriffe und Verfahren so erläutert, dass sie auch mit geringem Vorwissen verständlich sind.
- Darüber hinaus soll dieses Heft dazu beitragen, dass Sie nicht von den vielen Daten und Fachbegriffen „erschlagen“ werden, sondern wissen, was sich dahinter verbirgt und wo die Grenzen der Interpretation statistischer Kennwerte liegen.

Denn tatsächlich ist es so, dass Sie im Rahmen Ihrer Tätigkeit als Lehrerin oder Lehrer an vielen verschiedenen Stellen Ihrer beruflichen Tätigkeit immer wieder mit statistischen Auswertungen und Interpretationen statistischer Daten konfrontiert werden. Dazu zählen beispielsweise die Ergebnisse von nationalen und internationalen Leistungsvergleichsstudien (z. B. TIMSS, PISA und IGLU) sowie Rückmeldungen zu Vergleichsarbeiten bzw. Lernstandserhebungen und zentralen Abschlussarbeiten ebenso wie die Ergebnisse interner oder externer Evaluation (Schulinspektion) von Schulen. Um diese Daten für Ihre schulische Entwicklungsarbeit nutzen zu können, benötigen Sie gewisse statistische Grundkenntnisse. Die Ergebnisse statistischer Erhebungen und Analysen werden in zunehmendem Maße innerhalb des Kollegiums, der Eltern- und Schülerschaft sowie der zuständigen Aufsichtsbehörden diskutiert –



nicht zuletzt, um damit notwendige mittel- und langfristige Schulentwicklungsprozesse steuern zu können.

Die folgende Abbildung verdeutlicht die Anwendungsbereiche von Statistiken im Schulalltag.



**Abbildung 1:** Anwendung statistischer Kennwerte im Schulalltag (nach Perels; Steiner 2006)

## 2. Praxisbeispiele und deren Interpretation

Im Folgenden werden anhand von Praxisbeispielen wichtige statistische Kennwerte eingeführt und ihre Relevanz verdeutlicht.

### 2.1 Ergebnisse interner und externer Evaluation von Schule

Bei der **externen Evaluation** von Schule (Schulinspektion) werden in der Regel nicht nur die von der Schule bereitgestellten Schuldokumente analysiert, sondern auch Daten durch Interviews, Beobachtungen und Fragebögen erhoben, in denen subjektive Einschätzungen verschiedener Schulgruppen (Lehrerinnen und Lehrer, Eltern, Schülerinnen und Schüler) zu Schule, Unterricht und Schulgemeinde ermittelt werden. Im Inspektionsbericht werden die erhobenen Daten zum einen durch ausformulierte Beschreibungen im Fließtext und zum anderen durch quantitative Bewertungen ausgewählter Kriterien des Hessischen Referenzrahmens Schulqualität zurückgemeldet. Zusätzlich wird den Schulen eine CD zur Verfügung gestellt, auf der alle Rohdaten, d. h. die Beobachtungsdaten und die Ergebnisse der Fragebogenerhebung zusammengestellt sind. In diesem Zusammenhang werden auch statistische Itemkennwerte wie **Mittelwert** und **Streuung** genannt. Um die Gesamturteilsfindung für alle Beteiligten transparent zu machen, werden die Vorgehensweise und fast alle verwendeten Instrumente (z. B. Fragebogen, Unterrichtsbeobachtungsbogen) veröffentlicht und in dem überreichten Bericht wird das Zustandekommen der Ergebnisse erklärt. Dabei wird auch dokumentiert, wie viele Befragungen bzw. Beobachtungen in der Schule durchgeführt wurden (mit Bezug zur Grundgesamtheit) und zu welchen Ergebnissen diese Befragungen geführt haben.

Statistische  
Itemkennwerte:  
Mittelwert und  
Streuung



Für die Berechnung des Mittelwerts (arithmetisches Mittel) wird die Summe der einzelnen Messwerte gebildet und durch die Gesamtanzahl aller Werte dividiert.

Das gebräuchlichste Streuungsmaß ist die **Standardabweichung** (diese wird meist mit  $SD = \text{Standard Deviation}$  abgekürzt). Sie gibt Auskunft darüber, wie weit die Messwerte durchschnittlich um den Mittelwert liegen bzw. entfernt sind. Die Standardabweichung ist also ein Maß für die Unterschiedlichkeit der Einzelergebnisse.



Bei den durchgeführten Befragungen und Beobachtungen und besonders bei **schulintern erhobenen** Daten ist es wichtig, sich noch einmal die Durchführungsbedingungen der Befragung zu vergegenwärtigen. Bei der **Interpretation und Präsentation** sollten Sie insbesondere prüfen:

### Welche Gruppen wurden befragt und wie groß ist die zugrunde liegende Grundgesamtheit?

**Beispiel:** Eine interne Umfrage unter den Kollegen hat ergeben: „60 % der Befragten zeigen sich mit dem Informationsfluss zufrieden oder sogar sehr zufrieden“. Ihr Kollegium umfasst 30 Lehrkräfte (100 %).

Fragen zur Überprüfung der Durchführungsbedingungen einer Befragung

### Welcher Anteil der Grundgesamtheit wurde befragt?

Es stellt sich heraus, dass lediglich fünf der 30 Kollegen, also rund 17 % des Gesamtkollegiums, befragt werden konnten. Damit verbergen sich hinter den „60 % der Befragten“ nur drei Kollegen, also nur noch 10 % des Gesamtkollegiums.

### Wie wurden die Befragten (die Stichprobe) ausgewählt?

Bei den fünf befragten Kollegen handelt es sich, wenn auch nach Zufallsprinzip gezogen, um keine repräsentative Stichprobe (hierfür müssten in der getroffenen Auswahl alle im Kollegium vorhandenen, für die Umfrage relevanten Merkmale – z. B. Alter, Geschlecht, Dienstjahre, Deputat etc. – anteilmäßig repräsentiert sein). Besonders kritisch ist die Befragung dann, wenn dieser kleine Anteil nicht zufällig ausgewählt wurde, sondern es sich um (die einzigen) freiwillige(n) Teilnehmer handelt. Es haben möglicherweise nur hoch motivierte oder besonders (un-)zufriedene Lehrkräfte teilgenommen. Dadurch wird eine Übertragung der Ergebnisse auf die Gesamtheit der Lehrkräfte unmöglich.

### Welche Fragen wurden gestellt und welche Antwortkategorien standen zur Verfügung? Wie kommen also Mittelwerte zustande? Sind sich die Befragten in ihrer Meinung eher einig oder nicht?

**Beispiel:** Im Fragebogen gab es sechs Ankreuzmöglichkeiten, die höchste Zustimmung wurde nach Konvention mit dem höchsten Wert belegt (in diesem Fall 6, für „außerordentlich zufrieden“), die niedrigste Zustimmung mit dem niedrigsten Wert (1 für „überhaupt nicht zufrieden“). Die Mitte bzw. der Mittelwert unserer Beispiel-





skala (Abb. 2) liegt damit bei 3,5 – durch das rechnerische Verfahren ein Zahlenwert zwischen zwei Kästchen, der nirgendwo angekreuzt werden konnte.

Zustimmungsgrad:	3,5						gerundete Standardabweichung (SD) in Kästchen
	1	2	3	4	5	6	
Beispiel a	15					15	2,5
Beispiel b			15	15			0,5
Beispiel c	5	5	5	5	5	5	1,6
Beispiel d	3	8	4	4	8	3	1,7

**Abbildung 2:** Beispielverteilungen für 30 Befragte, Mittelwert 3,5 (schwarzer Balken); Beispiele a, b, c, d

**Welche Aussage machen Mittelwerte?**

Der Mittelwert allein sagt zunächst gar nichts über die tatsächliche Verteilung der Kreuzchen auf den Fragebogen aus, wie die Beispiele a bis d verdeutlichen. Obwohl die Kreuzchen in den Beispielen sehr unterschiedlich verteilt sind, ergibt sich nämlich derselbe Mittelwert. Daher ist es notwendig, das Streuungsmaß mit anzugeben. Dieses Streuungsmaß (z. B. die Standardabweichung) gibt dann z. B. die durchschnittliche Abweichung aller Teilnehmerinnen und Teilnehmer von dem berechneten Mittelwert an: In allen vier Beispielen ist der Mittelwert 3,5, die Standardabweichungen sind jedoch sehr unterschiedlich. Ist im Vergleich zur Gesamtlänge der Antwortskala die Streuung hoch (2,5), wie in Beispiel a (Abb. 2), liegen viele Werte weit vom Mittelwert entfernt. In Beispiel a liegen sogar alle Werte weitestmöglich, nämlich 2,5 Kästchen, vom Mittelwert entfernt, was auf eine Polarisierung der betroffenen Stichprobe hinsichtlich des Gegenstands der Frage hindeutet. Anders verteilte Meinungsbilder können Sie den Beispielen b, c und d entnehmen. In Beispiel b liegt eine sehr geringe Streuung (0,5) vor, da alle Antworten sehr nah beieinander liegen, d. h. die Einschätzungen der Befragten ähneln sich stark. Von den insgesamt sechs Antwortmöglichkeiten wurden nur zwei nebeneinanderliegende Kästchen angekreuzt. Anders verteilte Meinungsbilder können Sie den Beispielen c und d entnehmen – in diesen Fällen liegen die Streuungen zwischen den extremen Fällen von a und b.



### Ist die grafische Darstellung der Daten korrekt?

Bei der Rückmeldung von Daten durch Abbildungen treten immer wieder grundlegende Fehler oder verzerrte Darstellungen auf. Achten Sie bei der Interpretation der Abbildungen auf die Probleme grafischer Abbildungen, die in Abschnitt 3.3 beschrieben werden.

Für die grafische Darstellung von Daten können Balken- oder Säulendiagramme, Kreisdiagramme, Boxplots oder Linien-  
diagramme verwendet werden.

In den Abbildungen 3 und 4 werden zwei Balkendiagramme dargestellt, die aufgrund der gleichen Daten entstanden sind. Es geht dabei um die Frage, wie zufrieden die Lehrerinnen und Lehrer einer Schule mit dem Informationsfluss sind (s.o.). In der linken Abbildung sind die x- und y-Achsen beschriftet. Auch die Werteskala ist vollständig dargestellt. In der rechten Abbildung (Abb. 4) werden die gleichen Daten ohne Beschriftung und verzerrt dargestellt.

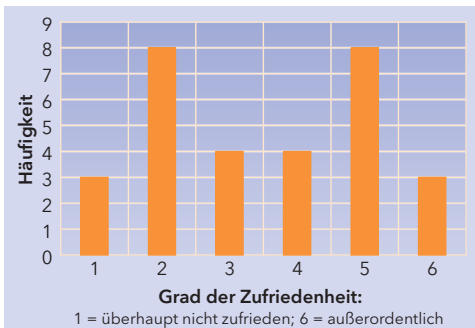


Abbildung 3: korrekte Darstellung

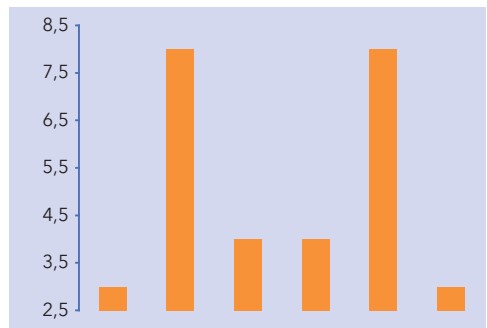


Abbildung 4: unvollständige und verzerrte Darstellung

## 2.2 Zentrale Lernstandserhebungen

Zentrale Lernstandserhebungen werden als nationale Erhebungen durchgeführt und in verschiedenen Jahrgangsstufen und Fächern geschrieben. Ziel ist es, regelmäßig den Lernstand von Schülerinnen und Schülern anhand landesweit einheitlicher Aufgaben festzustellen. Die Rückmeldung der Ergebnisse an die Schulen er-

### Adjustierung der Ergebnisse für die Rückmeldung

folgt hier in erster Linie anhand von Balkendiagrammen, seltener auch anhand von Boxplots (siehe Punkt 3.3) und Quartileinteilungen (siehe Punkt 3.2). Zum Teil werden die Werte für die Rückmeldungen – wie auch bei der PISA-Studie bekannt geworden – adjustiert, d. h. soziale Umfeldfaktoren (wie z. B. spezielle Charakteristika der Klassenzusammensetzung aufgrund des Anteils an Jungen und Mädchen, Klassenwiederholern, Schülerinnen und Schülern mit nicht deutschem Sprachhintergrund etc.) werden aus den Ergebnissen herausgerechnet. Das bedeutet, dass die Unterschiede zwischen Ländern, Schulen oder Klassen statistisch berücksichtigt werden und so Vergleiche vor dem Hintergrund „identischer“ Ausgangsbedingungen erstellt werden können. Nach der Adjustierung können demnach Faktoren wie z. B. ein hoher Anteil an Schülerinnen und Schülern mit Migrationshintergrund innerhalb einer Klasse als Erklärung von Unterschieden ausgeschlossen werden.

### Ergebnis- rückmeldung mit Hilfe von Quartilen

#### Rückmeldungen über Quartilierung aufgrund der Platzierung innerhalb der Rangreihe der Teilnehmer

Werden einzelne Schülerinnen und Schüler, Schulen oder Regionen in einer aufsteigenden Rangfolge bezüglich der erreichten Ergebnisse über- oder nebeneinander aufgelistet und wird diese Liste dann in vier Teile geteilt, erhält man Quartile. Die Quartilierung gibt einen Eindruck davon, wie Einzelleistungen von Personen im Vergleich zu den Leistungen aller teilnehmenden Personen einzuschätzen sind; sie sagt ohne weitere Angaben (z. B. Gesamtpunktzahl) allerdings nichts über die Wertigkeit der Leistung aus. Dies sollten Sie, falls Sie eine quartilierte Rückmeldung erhalten, berücksichtigen. Zwei Beispiele sollen dies verdeutlichen:

**Beispiel 1:** Sollten nur wenige Schulen einer Region sehr gute Leistungen erbracht haben, alle anderen aber sehr schlechte, könnte sich auch eine der Schulen mit schlechten Ergebnissen mit einem Platz im obersten Quartil rühmen, weil das oberste Viertel der teilnehmenden Schulen weit in den schlechten Leistungsbereich hineinreicht.

**Beispiel 2:** Befinden sich alle Schulen im obersten Leistungsbereich (vergleichbar der Schulnote „sehr gut“), können selbst Schulen mit Leistungen im obersten Leistungsbereich („sehr gut“) im untersten Quartil liegen.





**Vorteil:** Mit der Quartilierung wird der direkte Vergleich zwischen den Schulen entschärft. Sie gibt innerhalb des Quartils nur eine gemeinsame Lage in der gleichen Teilmenge, nicht das direkte Verhältnis zueinander an.

**Nachteil:** Die Art der Quartilierung kann für die innerhalb der Rangfolge der Teilnehmerinnen und Teilnehmer „schlechteren“ Personen nachteilig sein, vor allem, wenn diese in einem absolut betrachtet akzeptablen Leistungsbereich liegen.

Aus einer quartilierten Darstellung wird erkennbar, welche Werte unterhalb des ersten Quartils (25 %) und unterhalb des dritten Quartils (75 % der Werte) liegen. Dabei kann der Begriff Quartil sowohl für die Grenzen als auch für die Bereiche verwendet werden, die durch diese Einteilung entstehen.

### Rückmeldungen über Boxplots – die visualisierte Quartilierung

Eine weitere Form der Rückmeldung ist der Boxplot. Sie können dort auf einen Blick die Verteilung aller erreichten Werte in einer Grafik erschließen. Die Problematik der Interpretation von Mittelwerten ergibt sich hier nicht, da die Spannweite (Range) der erzielten Werte in der Grafik klar erkennbar ist (siehe Abbildung 8 und die genaueren Erläuterungen auf S. 20).

Boxplots zur Darstellung von Ergebnissen

Leitfragen zur Interpretation von Rückmeldungen im Rahmen von Lernstandserhebungen und Kompetenztests können sein:

- Wurden soziale Kontextfaktoren berücksichtigt?
- Wo liegen die Quartilsgrenzen und wie ist die objektive Leistung?
- Gibt es eine breite Streuung (beim Boxplot: lange Antennen, langer Körper), die ein Hinweis auf Leistungsheterogenität ist?

## 2.3 (Internationale) Schulleistungstudien

Während bei den nationalen Verfahren, wie z. B. den Lernstandserhebungen, jeweils die Grundgesamtheit (z. B. alle Schülerinnen und Schüler einer bestimmten Jahrgangsstufe) bekannt ist und in der Regel voll ausgeschöpft werden kann, kann dies für Leistungs-



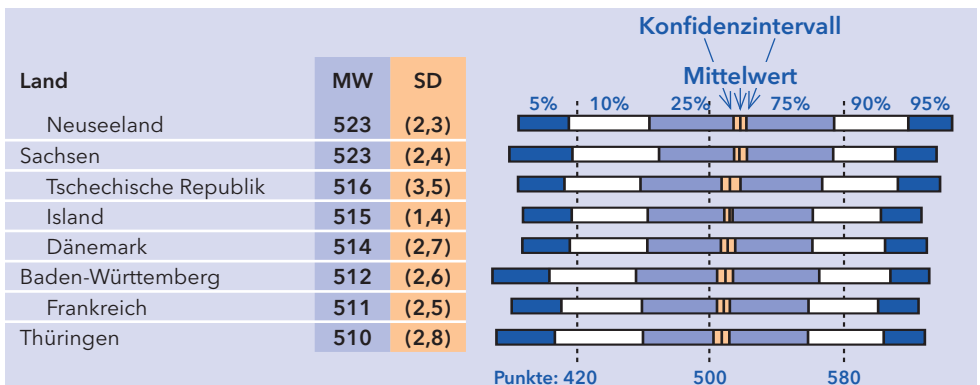


vergleichsstudien zum einen aus Gründen der Ökonomie, zum anderen aus Gründen der Zumutbarkeit nicht ermöglicht werden. Entsprechend muss ein größerer Aufwand bei der Stichprobenziehung betrieben werden, um innerhalb bestimmter Konfidenzintervalle verlässlich von der gezogenen Stichprobe (bei der PISA-Studie z. B. 15-jährige Schülerinnen und Schüler aus verschiedenen Jahrgangsstufen und Schultypen) auf die Gesamtpopulation (z. B. aller 15-Jährigen) zu schließen. Dann können, obwohl nur ein bestimmter Anteil der Gesamtpopulation der 15-Jährigen untersucht wurde, innerhalb bekannter Grenzen verlässliche Aussagen über die Gesamtpopulation gemacht werden, da statistisch berechnet werden kann, welche Werte in der Gesamtheit vorliegen.

### Rückmeldung mit Hilfe von Perzentilbändern

Die Rückmeldung der Pisa-Studie erfolgte in der Öffentlichkeit zu großen Teilen mit Hilfe von Perzentilbändern. Die dargestellten Perzentilwerte sind Maßzahlen, die darüber Auskunft geben, wie hoch der Kennwert für eine Merkmalsausprägung (z. B. Lesekompetenz) an bestimmten Abschnitten ist. Bei einem Perzentil (Prozentrang) von 95 % wird beispielsweise der Punktwert berichtet, ab dem die besten 5 % einer Verteilung beginnen, die Perzentilwerte 25 und 75 betreffen demnach das untere und obere Viertel einer Verteilung. Anhand der Perzentilwerte kann auch abgelesen werden, wie stark die Leistungsspitze (95 %-Perzentil) und wie schwach die untersten Gruppen (5 %-Perzentil) sind. Die folgende Abbildung soll dies noch einmal beispielhaft verdeutlichen:

**Abbildung 5:** Ausschnitt der Perzentilbänder aus der Pisa-Rückmeldung mit Konfidenzintervall (siehe dazu Abschnitt 3.4.2) um den Mittelwert



## 3. Schulrelevante statistische Kennwerte und Verfahren

Im Folgenden finden Sie detaillierte Beschreibungen der einzelnen statistischen Kennwerte, die im ersten Abschnitt nur jeweils kurz erwähnt wurden. Die Skalenniveaus bilden die Grundlage der erhobenen Daten, von ihnen hängt ab, welche statistischen Verfahren angemessen sind.

### 3.1 Skalenniveaus

Im schulischen Kontext wird eine große Anzahl von Daten erhoben. Bei diesen Messungen werden bestimmten Ausprägungen eines Merkmals (z. B. Alter, schulische Leistung usw.) bestimmte Zahlen zugeordnet (z. B. je älter eine Person, desto höher die ihr zugeordnete Zahl). Es lassen sich dabei verschiedene Ebenen (Skalenniveaus) unterscheiden, auf denen diese Messungen vollzogen werden können. Wichtig ist dabei, dass sich jedes Skalenniveau durch bestimmte Eigenschaften auszeichnet und deshalb nur bestimmte statistische Auswertungen zulässig sind.

Das niedrigste Skalenniveau, mit den geringsten Anforderungen an die Daten, wird als **Nominalskala** bezeichnet. Eine Nominalskala ermöglicht Aussagen über Gleichheit und Verschiedenheit von Objekten und leistet eine Gruppierung in Kategorien. Mit „Kategorien“ sind Einteilungen gemeint, die vorgenommen werden. Solche Einteilungen können naturgegeben sein (z. B. Geschlecht) oder auch bewusst gebildet werden (beispielsweise Eingruppierung der Schülerinnen und Schüler einer Schulklasse in schlechte, mittlere und gute Leser). Die einzelnen Zahlen, die den Kategorien oder Merkmalsausprägungen zugeordnet werden, sind beliebig wählbar. Ihnen kommt also keinerlei empirische Bedeutung zu, d.h. die Kategorien können anhand der zugeordneten Zahlen nicht interpretiert werden. Die Zahlen geben nur Auskunft darüber, ob zwei Objekte in eine Kategorie fallen oder zu unterschiedlichen Kategorien gehören (z. B. 1 = weiblich, 2 = männlich; hier kann nicht davon ausgegangen werden, dass die Zahl 2 für ein „größer“ oder „mehr“ steht). Typische Beispiele für Variablen, die Nominalskalenniveau besitzen, sind das Geschlecht oder die Einteilung, ob die Schüler an einer bestimmten Veranstaltung teilnehmen (z. B. 1 = keine Teilnahme am Wahlunterricht, 2 = Teilnahme am Wahlunterricht).

#### Nominalskala

### Ordinalskala

Von **Ordinalskalenniveau** wird gesprochen, wenn die Daten vergleichende Aussagen wie „größer/kleiner“ oder „besser/schlechter“ und „gleich/ungleich“ zulassen. In diese Kategorie fallen z. B. die Schulnoten. Dabei ist klar, dass die Note „1“ besser ist als die Note „2“. Es kann jedoch nicht gesagt werden, dass die Abstände zwischen zwei Notenstufen (z. B. zwischen „1“ und „2“ und zwischen „3“ und „4“) für die gleiche Leistungsdifferenz sprechen. Dies wird auch daran deutlich, dass die üblichen Prozentanteile der Gesamtpunktzahl, die für die einzelnen Noten erreicht werden müssen, nicht identisch sind.

### Intervallskala

Daten, die eine Differenzenbildung (z. B.  $A - B = C - D$ ), eine „größer/kleiner“- oder „besser/schlechter“-Relation und Aussagen über Unterschiede (gleich/ungleich) zulassen, haben **Intervallskalenniveau**. Typisch für dieses Messniveau sind z. B. Messungen der Temperatur in Grad Celsius. Verhältnisbildungen sind hier nicht erlaubt (heute  $14^{\circ}\text{C}$  sind nicht doppelt so warm wie gestern  $7^{\circ}\text{C}$ , sondern  $7^{\circ}\text{C}$  wärmer, da die Celsius-Skala keinen absoluten Nullpunkt als Bezugspunkt hat).

Erst bei diesem Skalenniveau ist es sinnvoll, den Mittelwert einer Variablen zu bestimmen. Für die meisten Fragebögen mit mehrstufigen Antwortskalen (z. B. 4-stufig mit „trifft nicht zu“, „trifft eher nicht zu“, „trifft eher zu“, „trifft zu“) wird ein solches Intervallskalenniveau angenommen. Dies ist nicht sinnvoll, wenn nur ein Nominalskalenniveau vorhanden ist (z. B. Geschlecht) oder Ordinalskalenniveau (z. B. Schulnoten, auch wenn die Bildung von Durchschnittsnoten eine gängige Praxis ist).

### Verhältnisskala

Das höchste Skalenniveau besitzt die **Verhältnisskala**, die einer Intervallskala entspricht, die einen denkbaren niedrigsten Wert enthält. Erst hier kann man auch Verhältnisse angeben (z. B. Längen oder Gewichte im Alltag, gestoppte Zeit). Bei den meisten Auswertungsverfahren kommt der Unterscheidung zwischen Intervallskala und Verhältnisskala weniger Bedeutung zu, denn für die Berechnung üblicher statistischer Kennwerte ist das Intervallskalenniveau ausreichend.

In Tabelle 3 werden die Zusammenhänge zwischen Skalenniveaus und den deskriptiven (beschreibenden) statistischen Kennwerten,



die für das jeweilige Niveau berechnet werden können, noch einmal zusammengefasst.

Skalenniveaus und statistische Kennwerte			
Skalenniveau	Empirische Aussage	Beispiel	Statistische Kennwerte
Nominal	gleich/ungleich	Geschlecht	- Häufigkeiten - Modalwert: häufigster Wert
Ordinal	gleich/ungleich, größer/kleiner bzw. besser/schlechter	Schulnoten	- (kumulierte) Häufigkeiten - Median: Wert, unterhalb und oberhalb dessen jeweils die Hälfte der Werte liegen
Intervall	gleich/ungleich, größer/kleiner bzw. besser/schlechter, Differenzbildung	Temperatur in Grad Celsius	- Mittelwert (arithmetisches Mittel): Summe der Messwerte geteilt durch ihre Anzahl - Standardabweichung/Varianz: (Quadrat der) Wurzel aus der Summe aller quadratischen Abweichungen der Messwerte vom Mittel geteilt durch die Fallanzahl - Quartile: unterhalb des 1. Quartils liegen 25 % der Werte und unterhalb des 3. Quartils liegen 75 % der Werte
Verhältnis	Interpretation von Verhältnissen	Längenmaße	- Einsetzbarkeit aller Kennwerte wie bei Intervallskalenniveau

**Tabelle 3:** Skalenniveaus und mögliche Berechnungen statistischer Kennwerte

### 3.2 Statistische Kennwerte

In diesem Abschnitt werden wichtige statistische Kennwerte, die sehr häufig für Auswertungen verwendet werden, näher erläutert. Dabei werden die deskriptiven Maße Modalwert, Median, Mittelwert, Standardabweichung, Varianz und Quartile erläutert. Wie bereits oben angedeutet, hat die Art des Skalenniveaus einen Einfluss darauf, welche Art der Auswertung für vorliegende Daten sinnvoll ist.



**Modalwert** Ein Modalwert beschreibt den am häufigsten vorkommenden Wert einer Verteilung. Modalwerte werden zur Beschreibung dann verwendet, wenn die Verteilung der Häufigkeiten ein Maximum besitzt und die Häufigkeit kleinerer und größerer Werte abfällt (siehe Tabelle 4). Für Daten mit Nominaldatenniveau ist das Zählen bzw. die Angabe der Häufigkeit für die einzelne Kategorie die einzige sinnvolle Auswertungsmöglichkeit.

	Messwert	Häufigkeit
	1	8
	2	18
Modalwert	3	21
	4	11
	5	3

**Tabelle 4:** Modalwert einer Nominalskala

**Median** Der Median beschreibt den Wert, unterhalb und oberhalb dessen jeweils die Hälfte der Werte liegen. Voraussetzung für die Berechnung des Medians ist Ordinalskalenniveau, d. h. größere Werte bedeuten eine stärkere Merkmalsausprägung. Wenn eine ungerade Anzahl an Fällen vorliegt, kann der Median bestimmt werden, indem die Messwerte der Größe nach geordnet und  $(n-1)/2$  Werte abgezählt werden. Der nächstgrößere Wert ist dann der Medianwert. Falls zum Beispiel eine Klasse aus 15 Schülern mit folgenden Noten besteht: 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 5; ist der Median der achte Wert  $((15-1)/2+1)$  und dieser ist eine 3. Ist die Gesamtzahl gerade, so werden die unteren und oberen 50 % der Fälle abgezählt und der Mittelwert zwischen dem größten Wert der unteren 50 % und dem kleinsten Wert der oberen 50 % gebildet. Kommt in dem Beispiel also noch ein Schüler mit der Note 2 hinzu, so ergibt sich folgender Median:



1, 1, 2, 2, 2, 2, 2, 2 | 3, 3, 3, 3, 3, 4, 4, 5  
50 %

Der Median setzt sich in diesem Fall zusammen aus  $(2+3)/2 = 2,5$ .



### Kumulierte Häufigkeiten

Für Ordinalzahlen können zusätzlich auch kumulierte Häufigkeiten angegeben werden. Kumuliert bedeutet hierbei „aufgehäuft“, d. h. die beobachteten Häufigkeiten werden schrittweise aufsummiert angegeben.

kumulierte  
Häufigkeit

Note	Häufigkeit	Kumulierte Häufigkeit	Rechnung
1	2	2	
2	6	8	$2 + 6 = 8$
3	5	13	$8 + 5 = 13$
4	2	15	$13 + 2 = 15$
5	1	16	$15 + 1 = 16$
6	0	16	

**Tabelle 5:** Beispiel für kumulierte Häufigkeiten

Kumulierte Häufigkeiten sind dann von Interesse, wenn nicht nur einzelne Häufigkeiten interessieren, sondern auch zusammengefasste Einheiten. So kann es beispielsweise interessant sein zu betrachten, wie viele Schülerinnen und Schüler die Note 3 oder besser erzielt haben. Für solche Fragestellungen geben kumulierte Häufigkeiten einen schnellen Überblick.

### Mittelwert

Der Mittelwert (arithmetisches Mittel) kann nur für intervallskalierte Daten sinnvoll berechnet werden. Hierbei wird die Summe der einzelnen Messwerte gebildet und durch die Gesamtanzahl aller Werte dividiert. Setzt sich beispielsweise eine siebte Klasse aus 25 Schülern mit folgendem Alter zusammen: 11 (1x); 12 (13x); 13 (8x) und 14 (3x), so wird die Gesamtsumme gebildet und durch 25 geteilt:  $(1 \times 11 + 13 \times 12 + 8 \times 13 + 3 \times 14) = 313$ ;  $313 / 25 = 12,52$ . Der Mittelwert wird häufig mit  $\bar{x}$  bezeichnet.

Mittelwert

Die mathematische Formel für die Berechnung des arithmetischen Mittels lautet:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

n = Stichprobengröße (z. B. Anzahl der untersuchten Schüler)



Es werden also für die Berechnung des Mittelwerts alle Werte (von 1 bis n) summiert und durch die Gesamtanzahl (n) dividiert.

**Streuungsmaß:  
Standard-  
abweichung**

Der Mittelwert ist nur dann sinnvoll interpretierbar, wenn zusätzlich Angaben über die Art der Verteilung der Messwerte gemacht werden. Das gebräuchlichste Dispersionsmaß (Streuungsmaß) ist die **Standardabweichung** (diese wird in Tabellen meist mit **SD = Standard Deviation** abgekürzt, in Formeln mit s). Die Standardabweichung gibt Auskunft darüber, wie weit die Messwerte durchschnittlich um den Mittelwert streuen (vgl. auch Abb. 2).

Den Rechengang beschreibt die folgende Formel:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

So kann beispielsweise eine Lehrkraft die mittlere Punktzahl ermitteln, die in einem von ihr durchgeführten Test erreicht wurde. Die Standardabweichung gibt Auskunft darüber, wie unterschiedlich die Schülerleistungen sind. Ist die Standardabweichung sehr groß, gibt es viele Schülerinnen und Schüler, die bedeutend besser bzw. schlechter als der Durchschnitt sind. Ist die Standardabweichung gering, erreichen die Schülerinnen und Schüler ähnliche Testergebnisse.

**Streuungsmaß:  
Varianz**

Ein weiteres bekanntes Streuungsmaß ist die **Varianz** ( $s^2$ ). Dabei handelt es sich um die quadrierte Standardabweichung. Das bedeutet, dass die Berechnung mit der Formel der Standardabweichung bis auf die Wurzel identisch ist, d. h. es entfällt der letzte Schritt und die Wurzel wird nicht berechnet. Die Varianz wird oftmals für Berechnungen in der schließenden (d. h. schlussfolgernden) Statistik verwendet. Analog zur Standardabweichung gibt die Varianz an, wie weit die Messwerte durchschnittlich um den Mittelwert streuen. Sie ist beispielsweise ein Maß dafür wie unterschiedlich Schülerleistungen bei einem durchgeführten Test sind.



Darüber hinaus werden oftmals **Quartile** berechnet: Dabei liegen unterhalb des ersten Quartils 25 % der Werte und unterhalb des dritten Quartils 75 % der Werte. Der Begriff Quartil kann sowohl für die Grenze als auch für die durch diese Einteilung entstehenden Bereiche verwendet werden (siehe auch das Thema Boxplots). Quartile ähneln dabei den Perzentilen. Quartile haben jedoch immer fest definierte Grenzen (25 %, 50 % und 75 %), wohingegen Perzentilgrenzen auch andere Werte (z. B. 5 % und 95 %) annehmen können.

### 3.3 Darstellungsformen

Die Ergebnisse statistischer Untersuchungen werden auf vielfältige Art und Weise zurückgemeldet. Dabei werden häufig neben der Angabe der ermittelten statistischen Kennwerte (z. B. Mittelwerte und Standardabweichung) auch grafische Darstellungen der Ergebnisse verwendet, da durch dieses Vorgehen die Informationen anschaulicher werden. Im Folgenden werden die häufigsten grafischen Darstellungsarten vorgestellt: Balkendiagramme (Abbildung 6), Kreisdiagramme (Abbildung 7), Boxplots (Abbildung 8) und Liniendiagramme (Abbildung 9).

#### Balken- /Säulendiagramme

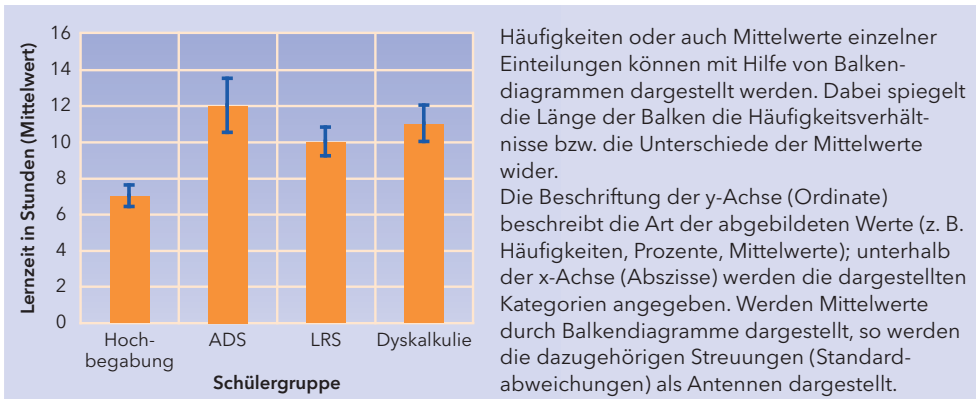
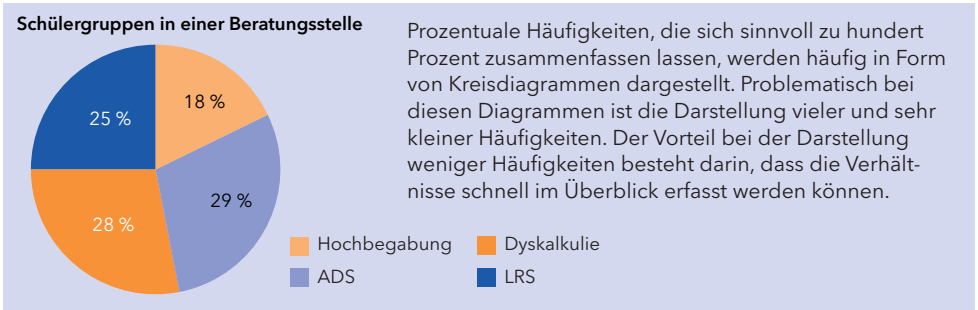


Abbildung 6: Beispiel eines Balkendiagramms

## Kreisdiagramme

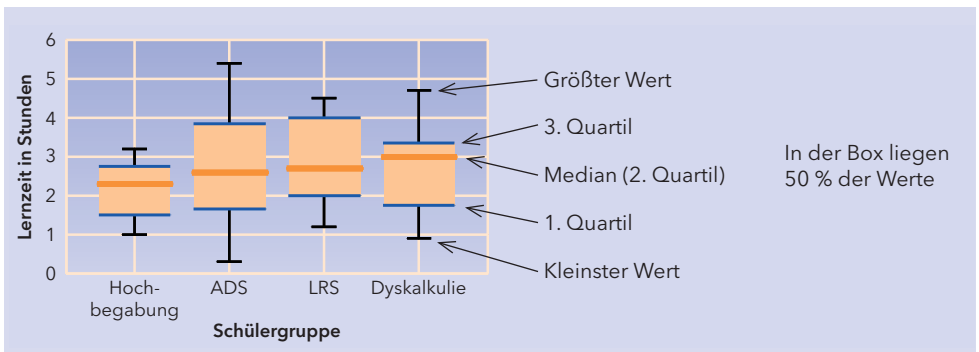


**Abbildung 7:**  
Beispiel eines  
Kreisdiagramms

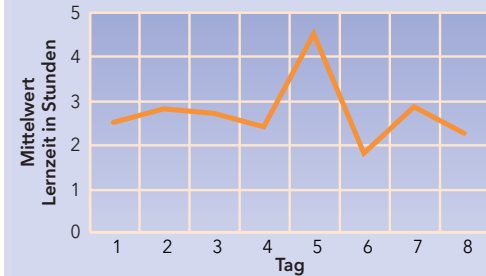
## Boxplots

Boxplots sind sehr beliebt, wenn die Quartile intervallskalierter Variablen dargestellt werden sollen. Variablen sind veränderliche Größen, die betrachtet werden, es kann sich um Alter, Testergebnisse, sozioökonomischen Status usw. handeln. Bei der folgenden Abbildung markieren die äußeren Linien (die Enden der Antennen) den kleinsten bzw. größten Wert. Die eine Begrenzung der Box ist das erste Quartil, die mittlere Linie markiert das zweite Quartil (Median) und die andere Begrenzung der Box kennzeichnet das dritte Quartil. In der Box des Boxplots liegen also insgesamt 50 % der Werte: 25 % der Werte bzw. der Teilnehmerinnen und Teilnehmer auf der einen und 25 % auf der anderen Seite des Medians. Durch diese Art der Darstellung werden alle Teilnehmerergebnisse und deren Verteilung in der Grafik erschließbar.

**Abbildung 8:**  
Beispiel eines  
Boxplots



## Liniendiagramme



Liniendiagramme werden häufig dann eingesetzt, wenn ein zeitlicher Verlauf dargestellt werden soll, d. h. wenn z. B. die gleiche Gruppe mehrmals befragt wurde. An der x-Achse (Abszisse) werden dann normalerweise die einzelnen Erhebungszeitpunkte (z. B. Tage) abgetragen.

## 3.4 Statistische Bedeutsamkeit: Signifikanz

Neben deskriptiven Analysen, also der beschreibenden Darstellung von Daten (z. B. durch Mittelwerte, Häufigkeitsverteilungen usw.), sind für viele Fragestellungen auch inferenzstatistische Aussagen von Bedeutung. Inferenzstatistik bedeutet schließende (d. h. schlussfolgernde) Statistik und beschäftigt sich mit der Fragestellung, wie aufgrund von Ergebnissen, die anhand einer (meist relativ kleinen) Stichprobe gewonnen wurden, allgemeingültige (generalisierbare) Aussagen abgeleitet werden können. So wird z. B. aufgrund der Ergebnisse einer stichprobenhaften Schülererhebung bei PISA auf die gesamte Schülerschaft eines Jahres geschlossen. Im Rahmen der Signifikanzprüfung kann entschieden werden, ob gefundene Unterschiede oder Zusammenhänge in einer Stichprobe nur zufällig zustande gekommen sind oder mit größerer Wahrscheinlichkeit (95 % oder 99 %) tatsächlich vorhanden sind. Deshalb kommt solchen Verfahren eine entscheidende Bedeutung zu.

### 3.4.1 Stichprobenarten

Von zentraler Bedeutung ist dabei die Art, wie die Stichprobe ausgewählt wird. Ideal ist es, eine Stichprobe zu ziehen, die in Bezug auf die relevanten zu untersuchenden Merkmale repräsentativ für die Grundgesamtheit (alle Personen, die hinsichtlich der Fragestellungen prinzipiell untersuchbar sind) ist, auf die geschlossen werden soll. Ist dies der Fall, wird von **geschichteten/stratifizierten**

### Abbildung 9:

Beispiel eines Liniendiagramms

**geschichtete/  
stratifizierte  
Stichproben**

## Zufalls- und Klumpenstichproben

**Stichproben** gesprochen. Da dies in den meisten Untersuchungen nicht möglich oder sinnvoll ist, werden häufig **Zufallsstichproben** gezogen. Die Schülerinnen und Schüler einer Schule werden zufällig für eine Befragung ausgewählt – jeder Schüler muss dabei mit der gleichen Wahrscheinlichkeit ausgewählt werden können (bezogen auf das Beispiel ist es wichtig, dass alle Schüler erfasst sind und auch nicht mehrmals aufgeführt sind, damit die Wahrscheinlichkeit, ausgewählt zu werden, für alle Schülerinnen und Schüler gleich ist). Darüber hinaus können auch **Klumpenstichproben** gezogen werden. Dies ist beispielsweise dann der Fall, wenn mehrere Klumpen (beispielsweise Schulklassen) zufällig ausgewählt und komplett (alle Schüler der Klassen) an der Untersuchung teilnehmen.

### 3.4.2 Konfidenzintervalle

#### Konfidenzintervalle

Das Ziel inferenzstatistischer Messungen besteht darin, aus der gezogenen Stichprobe (z. B. zufällig ausgewählte 15-jährige Schülerinnen und Schüler) Schlüsse auf die gesamte Verteilung (z. B. alle 15-jährigen Schülerinnen und Schüler in Hessen) zu ziehen und die Aussagen verallgemeinern zu können. In diesem Zusammenhang ist das **Konfidenzintervall** von Bedeutung. Das Konfidenzintervall ist der Wertebereich, in dem sich 95 % oder 99 % aller möglichen Werte der gesamten Population befinden, die den Wert in der vorliegenden Stichprobe erzeugt haben können (vgl. Bortz 2005, S. 101ff). Das bedeutet, dass im Konfidenzintervall der Stichprobe mit 95- oder 99-prozentiger Wahrscheinlichkeit der Wert der Gesamtpopulation liegt.

Erreicht beispielsweise eine Schulklasse ( $n = 33$ ) bei einem Lesetest im Durchschnitt einen Wert von 20,0 und liegt das 95-prozentige Konfidenzintervall zwischen 18,82 und 21,18, so liegt der Mittelwert der Grundgesamtheit (der Populationsparameter) mit 95-prozentiger Wahrscheinlichkeit zwischen 18,82 und 21,18.

### 3.4.3 Signifikanzniveau

Eine zentrale Aufgabe der inferenzstatistischen Verfahren ist es, mit Hilfe eines geeigneten statistischen Tests zu überprüfen, ob ermit-



telte Unterschiede zwischen zwei oder mehreren Gruppen rein zufällig entstanden sind oder nicht. Da auch in diesem Fall von einer Stichprobe auf eine Grundgesamtheit geschlossen wird, können diese Aussagen nicht mit 100-prozentiger Sicherheit getroffen werden. Im Allgemeinen ist es (sehr) zufriedenstellend, wenn die Wahrscheinlichkeit, eine fehlerhafte Entscheidung zu treffen, maximal 5 % oder 1% beträgt. Diese Wahrscheinlichkeit wird auch Irrtumswahrscheinlichkeit genannt und die 5 %- oder 1 %-Grenze wird auch als Signifikanzniveau bezeichnet. Ergebnisse, die eine Irrtumswahrscheinlichkeit von < 5 % aufweisen, gelten als signifikant (statistisch bedeutsam) und Ergebnisse mit einer Irrtumswahrscheinlichkeit von < 1 % als hoch signifikant.

## Signifikanzniveaus

### 3.5 Zusammenhangsmaße

Soll der Zusammenhang zwischen zwei Variablen (z. B. der Zusammenhang zwischen Intelligenzquotient und Mathematiknote) analysiert werden, so kann das mit Hilfe des **Korrelationskoeffizienten** geschehen. Dabei geht es um solche Zusammenhänge, die durch die Relationen „je größer Variable 1, desto größer Variable 2“ oder „je größer Variable 1, desto kleiner Variable 2“ beschrieben werden können (z. B. „je größer eine Person, desto schwerer ist sie“). Der Korrelationskoeffizient wird durch den Buchstaben  $r$  beschrieben. Sein Wertebereich liegt zwischen  $-1 < r < +1$ . Für die Größe/Stärke des Zusammenhangs der beiden Variablen ist allein der Betrag des Korrelationskoeffizienten von Bedeutung. Das Vorzeichen gibt an, ob es sich um einen gleichläufigen Zusammenhang (positives  $r$ : „je mehr, desto mehr“ bzw. „je weniger, desto weniger“, z. B. „je größer, desto schwerer“ bzw. „je kleiner, desto leichter“) oder um einen gegenläufigen Zusammenhang handelt (negatives  $r$ : „je mehr, desto weniger“ bzw. „je weniger, desto mehr“, z. B. „je vitaminreicher die Ernährung, desto geringer das Herzinfarktrisiko“ und „je weniger vitaminreich, desto größer das Risiko“). Die Korrelation lässt keine Kausalaussagen zu, d. h. über die Wirkrichtung des Zusammenhangs kann keine Aussage gemacht werden. (Zwar kann z. B. die Aussage getroffen werden: „Je größer ein Mensch, desto schwerer ist er.“ Es kann aber aufgrund eines Korrelationsergebnisses nicht geschlossen werden, dass der Mensch so schwer ist, weil er so groß ist.)





So wurde in vielen Untersuchungen ein positiver Zusammenhang zwischen sozioökonomischem Status (SES) und Schulleistung ermittelt. Allein durch diese Korrelationen lässt sich jedoch nicht klären, ob eines der Phänomene ursächlich für das andere ist – oder umgekehrt. Es ist also unklar, ob ein hoher SES die Ursache für eine gute Schulleistung ist oder umgekehrt eine gute Schulleistung für einen hohen SES ursächlich ist.

Korrelationskoeffizienten sind ein Zusammenhangsmaß zwischen zwei Variablen. Die Werte können zwischen  $-1$  und  $+1$  variieren. Falls die Korrelation einen Wert nahe Null besitzt, besteht kein systematischer Zusammenhang zwischen den untersuchten Variablen.

### 3.6 Grenzen und Probleme

#### Durchführung und Interpretation statistischer Auswertungen

Im Folgenden werden einige Probleme aufgezeigt, die sich im Rahmen der Durchführung und Interpretation von statistischen Auswertungen ergeben können und die Hinweise darauf geben, was in diesem Zusammenhang besonders beachtet werden soll. Dabei werden die Probleme den Themen zugeordnet, die im vorigen Abschnitt dargestellt wurden. Im Anschluss daran werden noch allgemeine Hinweise gegeben.

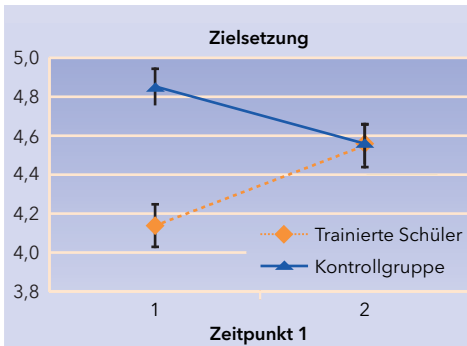
**Skalenniveau:** Bei der Analyse und Interpretation empirischer Daten muss immer das Skalenniveau beachtet werden. So ist es nicht zulässig und sinnvoll, Mittelwerte und Streuungen bei Daten zu berechnen, deren Niveau unterhalb des Intervalldatenniveaus liegt (z. B. Geschlecht). Das gilt auch für Schulnoten, die Ordinalzahlenniveau besitzen (so ist beispielsweise – obwohl dies im Schulalltag häufig geschieht – die Ermittlung von Durchschnittsnoten einer Klassenarbeit nicht sinnvoll).

**Mittelwerte/Streuungen:** Für die Interpretation statistischer Ergebnisse ist auch von Bedeutung, die Verteilung der Messwerte zu betrachten. Zusätzlich zum Mittelwert sollte daher immer ein Streuungsmaß zur Interpretation herangezogen werden, da dies Aufschluss über die Verteilung der Messwerte gibt.

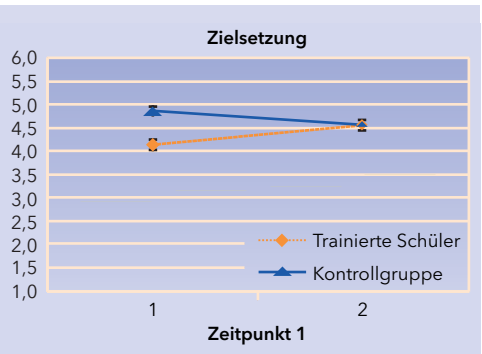


## Grafische Darstellung

Bei grafischen Darstellungen im Koordinatensystem sollte darauf geachtet werden, dass die Skalierung beider Achsen am niedrigsten Punkt der Skalierung beginnt und die Skalierung gleichmäßig ist, also beispielsweise der Abstand zwischen 1 und 2 mit dem zwischen 3 und 4 identisch ist. Ansonsten kann es zu verzerrten Eindrücken kommen. Abbildung 10 (unvollständige Skala) und Abbildung 11 (vollständige Skala) stellen die gleichen Werte dar – jedoch mit unterschiedlichen Skalierungen:



**Abbildung 10:** unvollständige Skala (da der Wertebereich von 1 bis 6 geht, aber nur von 3,8 bis 5 dargestellt ist)



**Abbildung 11:** Werte mit vollständiger Skala

**Zusammenhangsmaß – Korrelation:** Korrelationen geben einen Hinweis auf den Zusammenhang zweier Variablen („je mehr, desto mehr“ bzw. „je mehr, desto weniger“). Kausale Schlüsse, d. h. Aussagen über die Wirkrichtung, können und dürfen anhand von Korrelationen nicht gezogen werden. Dazu bedarf es theoretisch fundierter und klar belegter kausaler Wirkungszusammenhänge.

**Bedeutsamkeit statistischer Aussagen – Signifikanz:** Von statistisch signifikanten Aussagen bzw. statistisch abgesicherten Ergebnissen kann nur dann gesprochen werden, wenn tatsächlich unter Berücksichtigung aller Voraussetzungen (z. B. Skalenniveau, Stichprobenziehung) Signifikanztestungen vorgenommen und dokumentiert wurden. Ansonsten ist es möglich, dass vorliegende Ergebnisse



nur zufallsbedingt zustande gekommen sind und nicht generalisiert werden dürfen.

### Abschließende Hinweise

#### Allgemeine Hinweise:

Bei der Interpretation deskriptiver Analysen sollte immer wieder auf die befragte Gruppe Bezug genommen werden. Verallgemeinerungen auf die Grundgesamtheit sind nicht zulässig, sofern die Ergebnisse nicht einer repräsentativen Erhebung entstammen bzw. statistisch abgesichert sind.

Werden Prozentwerte angegeben, so sollte darauf geachtet werden, wie groß die absolute Zahl der Untersuchten ist. So könnten sich hinter einem Prozentsatz von 25 % eine Person (wenn insgesamt 4 Personen befragt wurden) oder auch 100 Personen (wenn insgesamt 400 Personen befragt wurden) verbergen.

Bei Ergebnissen, die aus Befragungen (Fragebogen) stammen, muss – soweit nicht nur Fakten (z. B. Schülerzahlen, Wochenstunden) erhoben wurden – deutlich werden, dass es sich dabei um subjektive Einschätzungen der Befragten handelt, die keinen Anspruch auf Objektivität haben. In diesen Fällen gibt es auch keine richtigen oder falschen Antworten.

In jeder empirischen Untersuchung, in welcher statistische Ergebnisse beschrieben werden, sollte auch die Art des Vorgehens (vor allem die verwendeten Messinstrumente) dargestellt werden. Ansonsten ist der Ablauf der Untersuchung und die Generierung der Ergebnisse nicht nachvollziehbar und es können (auch unbeabsichtigt) Ergebnisse verzerrt dargestellt werden.



## 4. Fazit

Mit diesem IQ Kompakt wollten wir vor dem Hintergrund der Anforderungen, vor die Sie im Rahmen der Tätigkeit als Lehrerin bzw. Lehrer gestellt werden, grundlegende Begriffe, Konzepte und Verfahren statistischen empirischen Arbeitens vorstellen und anhand von für Sie typischen Fragestellungen vertiefen. Der Schwerpunkt lag dabei auf der Interpretation deskriptiver Daten, da diese in der Schulpraxis von größerer Bedeutung sind.

Der Beitrag sollte Hinweise zum Lesen und Interpretieren von Tabellen geben und dabei auch auf statistische und testtheoretische Grundlagen eingehen. Bezogen auf inferenzstatistische Auswertungen wurde nur auf das Konfidenzintervall und die Bedeutung von Signifikanzaussagen eingegangen. Diese sind vor allem in Bezug auf die internationalen Vergleichsstudien von Bedeutung.

Wir hoffen, dass wir Ihrem persönlichen Anliegen gerecht werden konnten und Sie nach der Lektüre über eine Grundvorstellung von Statistik und Evaluation verfügen. Sie können die Broschüre auch immer wieder zur Hand nehmen, falls Sie statistische Darstellungen aus verschiedenen Bereichen des Schulwesens, beispielsweise Ergebnisse aus Lernstandserhebungen, interpretieren möchten. Schließlich können Sie die Broschüre als Nachschlagewerk für einzelne statistische Begriffe nutzen.

### Literaturangaben:

**Bortz, J.:** Statistik für Human- und Sozialwissenschaftler (6. Aufl.). Heidelberg 2005.

**Bortz, J.; Döring, N.:** Forschungsmethoden und Evaluation (4. Aufl.). Heidelberg 2006.

**Perels, F.; Steiner, I.:** „Traue keiner Statistik...“. Einführung in Statistik für Schulleiter. Schulleitung und Schulentwicklung. Erfahrungen – Konzepte – Strategien, 28, E1.2, 1-14. Berlin 2006.



## Warum ein IQ Kompakt zum Thema „Statistik und Evaluation“?

Das ist eine Frage, die Ihnen vielleicht durch den Kopf geht, wenn Sie dieses Heft in die Hand bekommen. Ziel dieses IQ Kompakt ist es, Ihnen die Grundbegriffe der Statistik und Evaluation so näher zu bringen, dass Sie die statistischen Daten und Auswertungsarten, mit denen Sie in Ihrem Schulalltag konfrontiert werden, verstehen und interpretieren können. Sie können es zur Hand nehmen, wenn Sie eine Lesehilfe benötigen oder wenn Sie statistische Begriffe nachschlagen möchten.

In dem Heft werden Ihnen zunächst einige Beispiele aus der Praxis vorgestellt, bei denen Sie zum Verständnis Statistik- und Evaluationskenntnisse benötigen. Dies ist bei den internationalen Vergleichsstudien (PISA, IGLU...) ebenso der Fall wie bei der Interpretation der Ergebnisse der nationalen und internationalen Vergleichsarbeiten bzw. Lernstandserhebungen sowie der zentralen Abschlussarbeiten. Danach wird ein Überblick über statistische Kennwerte und Verfahren gegeben, die im Schulkontext relevant sind.

Die Texte sind praxisnah anhand von Beispielen aus dem Schulalltag aufbereitet, und ermöglichen so einen guten Einblick in das Thema.

HESSEN



Institut für  
Qualitätsentwicklung

Institut für Qualitätsentwicklung (IQ)  
Walter-Hallstein-Str. 5-7  
65197 Wiesbaden  
[www.iq.hessen.de](http://www.iq.hessen.de)

