

Beschreibende Statistik

Silke Ruwisch

Ziel der beschreibenden Statistik ist die Erhebung, Aufbereitung und Interpretation von Daten. Der Beitrag vermittelt wichtiges Grundwissen – für den Unterricht und für den eigenen Alltag.

Die Behandlung des Themas „Daten“ im Mathematikunterricht dient vor allem der Mündigkeit im alltäglichen Handeln. Darüber hinaus erschließt sich damit ein Teilgebiet der Mathematik, welches lange innerhalb der Disziplin eher als Hilfsmittel, denn als eigenes Gebiet angesehen wurde: die elementare Statistik. Grob lässt diese sich in zwei Bereiche gliedern: die beschreibende oder deskriptive Statistik, deren Ziel die Erhebung, Aufbereitung und Interpretation von Daten zu bestimmten Fragestellungen ist, sowie die beurteilende oder induktive Statistik, deren Ziel das Aufstellen, Analysieren und Auswerten von Hypothesen ist.

► Merkmalsausprägungen

In der beschreibenden Statistik stehen „Daten“ im Mittelpunkt des Interesses. Ausgehend von einer Fragestellung werden die Elemente einer Menge hinsichtlich eines spezifischen Merkmals betrachtet. Die unterschiedlichen Werte, die jedem Element bezüglich dieses Merkmals zugeordnet werden, bezeichnet man als Merkmalsausprägungen (s. auch **KASTEN** auf S. 5). Die Festlegung möglicher Merkmalsausprägungen mag auf den ersten Blick banal erscheinen, erweist sich aber schon bei Aufgabenstellungen im Grundschulalter als diskussionswürdig und wichtig. Die möglichen Ergebnisse beim Würfeln mit einem Spielwürfel festlegen, das können bereits Erstklässler. Schon bei der Erhebung von Augen- oder Haarfarben sind jedoch sowohl bei der Vorgabe möglicher Antwortkategorien als auch bei der anschließenden Datenerhebung Interpretationen vorzunehmen. Wie viele Augenfarben sollen unterschieden werden? Blau, Grün, Braun und „sonstige“? Und wenn ein Kind grün-braune Augen hat, wird es zu den grün- oder zu den braunäugigen Kindern gezählt? Oder gehört es in die Kategorie „sonstige“? – Sicherlich lassen sich Kategorien immer weiter ausdifferenzieren, z. B. ließen sich die Kategorien Grün-Blau, Grau-Blau etc. aufnehmen. Gibt es jedoch letztlich zu jedem Merkmalsträger eine eigene Merkmalsausprägung, dann sind keine übergreifenden,

ABSOLUTE UND RELATIVE HÄUFIGKEIT

Angenommen, es gibt 10 blonde Kinder in einer Klasse von insgesamt 30 Schülerinnen und Schülern.

Die Merkmalsträger sind in diesem Fall die Kinder, das Merkmal ist die Haarfarbe und eine mögliche Merkmalsausprägung ist Blond (s. auch **KASTEN** auf S. 5).

Die absolute Häufigkeit der Kinder mit blonder Haarfarbe beträgt 10, die relative Häufigkeit ist gleich $\frac{10}{30}$, also ein Drittel.

Absolute Häufigkeit

Die absolute Häufigkeit ist die Anzahl der Merkmalsträger, die einer bestimmten Merkmalsausprägung zugeordnet sind. Die absolute Häufigkeit eines Messwerts gibt also an, wie oft dieser bestimmte Messwert innerhalb einer Stichprobe vorkommt.

Relative Häufigkeit

Die relative Häufigkeit gibt den Anteil der Merkmalsträger einer Merkmalsausprägung in Bezug auf die Gesamtanzahl der Merkmalsträger an. Die relative Häufigkeit eines Messwerts besagt also, wie oft dieser im Verhältnis zu allen Messwerten innerhalb einer Stichprobe vorkommt.

	QUALITATIVE MERKMALE		QUANTITATIVE MERKMALE	
	Nominalskala Merkmals- ausprägungen, die sich unterscheiden lassen	Ordinalskala Merkmals- ausprägungen, die sich in eine Reihe bringen lassen	Intervallskala Merkmals- ausprägungen, die gleiche Abstände haben	Verhältnisskala Merkmals- ausprägungen, die gleiche Abstände haben und denen ein absoluter Null- punkt zugrunde liegt
Mögliche mathematische Operationen		▶ Kleiner/Größer- Relation	▶ Kleiner/Größer- Relation ▶ Addition/ Subtraktion	▶ Kleiner/Größer- Relation ▶ Addition/ Subtraktion ▶ Multiplikation/ Division
Beispiele	▶ Geschlecht ▶ Haarfarbe ▶ Nationalität ▶ Automarke	▶ Noten ▶ Klasse(nstufe) ▶ Kleidergrößen ▶ Gehaltsstufen ▶ Handelsklassen	▶ Temperatur ▶ Intelligenz- quotient ▶ Zeit(punkte)	▶ Körpergröße ▶ Alter ▶ Zeit(spanne) ▶ Einkommen ▶ Wachstum
Mögliche Aussage	Es gibt blonde, braunhaarige und schwarzhaarige Kinder in unserer Klasse.	Für ein Kleid in Größe 36 braucht man weniger Stoff als für dasselbe Kleid in Größe 42.	Es ist heute 5° C wärmer als gestern.	Du bist mehr als doppelt so groß wie bei deiner Geburt.
Nicht mögliche Aussagen	Blond ist besser als braunhaarig.	Für ein Kleid in Größe 42 braucht man ein Sechstel mehr Stoff als für ein Kleid in Größe 36.	Es ist heute dreimal so warm wie gestern.	
Statistisch sinn- volle Kennwerte (s. auch KASTEN auf S. 25)	▶ Modalwert ▶ absolute/relative Häufigkeit	▶ Modalwert ▶ absolute/relative Häufigkeit ▶ Median	▶ Modalwert ▶ absolute/relative Häufigkeit ▶ Median ▶ arithmetisches Mittel	▶ alle statistischen Kennwerte

1 Qualitative und quantitative Merkmale

zusammenfassenden Aussagen möglich. Und gerade die sind ja das Ziel statistischer Erhebungen.

► Datentypen

Merkmale können qualitativer oder quantitativer Art sein. Je nach Typ lassen sich verschiedene mathematische Methoden anwenden und sind bestimmte Darstellungen passend **1**.

Qualitative Merkmale

Qualitative Merkmale, wie Geschlecht, Haar- und Augenfarbe, Lieblingstier, Staatsangehörigkeit, Kleidergröße oder Schulnote, lassen sich nur erfassen und auflisten, eventuell in Kategorien zusammenfassen, aber nicht mathematisch verrechnen. Allerdings lassen sich zwei Typen qualitativer Merkmale unterscheiden: Im Gegen-

satz zu den sogenannten nominalskalierten Daten lassen sich rang- oder ordinalskalierte Daten in eine Reihung bringen. So können z. B. Schulnoten oder Kleidergrößen der Größe nach geordnet werden und es lassen sich Aussagen treffen wie „Röcke der Größe 36 sind enger als Röcke der Größe 42.“ oder „Kinder, die ein ‚sehr gut‘ geschrieben haben, haben bessere Leistungen erbracht als diejenigen, die ein ‚befriedigend‘ erhalten haben.“ Eine derartige Reihung ergibt bei nominalskalierten Merkmalen wie Geschlecht oder Staatsangehörigkeit keinen Sinn, weil kein „mehr“ oder „weniger“ einer Merkmalsausprägung gegeben ist. Auch ordinalskalierte Merkmale lassen jedoch redlicherweise keine weiteren mathematischen Operationen zu: So kann beispielsweise nicht davon ausgegangen werden, dass der Leistungsunterschied zwischen 1 und 2 derselbe ist wie zwischen 4 und 5, sodass sich das Bilden des arithmetischen Mittels, also das Berechnen des Notendurchschnitts, verbietet.

Sind qualitative Daten in Diagrammen dargestellt, handelt es sich aufgrund der genannten Eigenschaften immer um die absoluten oder relativen Häufigkeiten (s. **KASTEN** auf S. 40). Die einzelnen Merkmalsausprägungen können selbstverständlich weiter zusammengefasst und die Häufigkeiten entsprechend addiert werden. Verschiedene Staatsangehörigkeiten ließen sich beispielsweise zusammenfassen zu denjenigen innerhalb der EU und denjenigen außerhalb oder zu denjenigen der verschiedenen Kontinente. Welche Gruppierung geeignet ist, ist abhängig von der Aussage, die verdeutlicht werden soll. Auch Aussagen wie „Mehr als die Hälfte der Kinder hat eine Note erhalten, die besser ist als befriedigend.“ lassen sich aus einer gruppierten Darstellung mit einem Blick erfassen. Werden die Daten dagegen nicht gruppiert dargestellt, müssen die einzelnen Zahlen noch addiert werden, um zu der obigen Aussage zu gelangen.

Üblicherweise werden die Kategorien bei nominalskalierten Daten hinsichtlich der Häufigkeit ihres Auftretens geordnet, während ordinalskalierte Daten eher in ihrer „natürlichen“ Reihenfolge belassen werden.

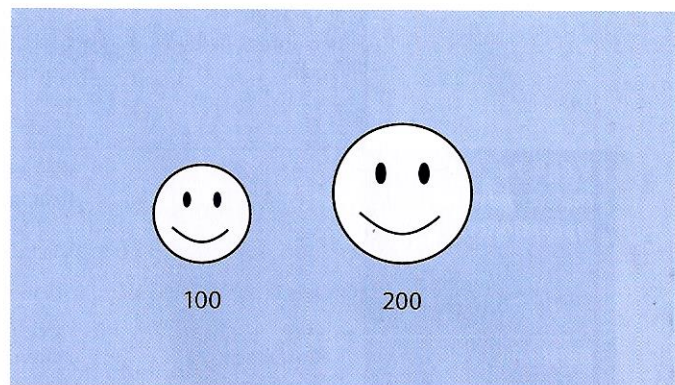
Quantitative Daten

Quantitativen Daten, wie z. B. Körpergröße, Lebensalter, Einkommen oder Temperatur, unterliegt im Gegensatz zu qualitativen Daten eine metrische Skala, d. h. Merkmalsausprägungen sind messbar und lassen sich mittels reeller Zahlen angeben.

Auch bei den quantitativen Daten lassen sich zwei Typen unterscheiden: intervall- und verhältnisskalierte Daten. Sie unterscheiden sich hinsichtlich des Nullpunktes der zugrunde liegenden Skala. Während z. B. bei unserem Thermometer der Nullpunkt 0°C (als Gefrierpunkt von Wasser) vereinbart wurde, kann man bei der Längenmessung von einem absoluten Nullpunkt ausgehen. Der Wert „Null“ bedeutet hier immer, dass keine Länge gegeben ist, unabhängig davon, ob mit Zollstock, Lineal oder Maßband, in Meter, Ellen oder Meilen gemessen wurde.

Daten, denen eine willkürliche Nullpunktsetzung unterliegt, heißen intervallskalierte Daten. Diese können addiert und subtrahiert werden, jedoch nicht multipliziert oder dividiert. Eine Aussage wie „Heute ist es 3°C wärmer als gestern.“ ist möglich, nicht jedoch die Aussage „Heute ist es dreimal so warm wie gestern.“

Auf verhältnisskalierte Daten, denen ein absoluter Nullpunkt unterliegt, lassen sich alle vier Rechenoperationen



2 Doppelte Anzahl – doppelte Fläche

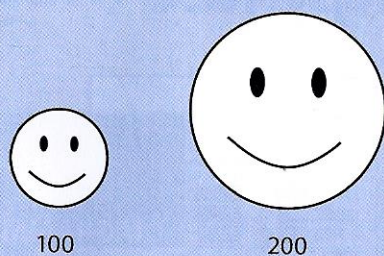
anwenden, also auch die Multiplikation und Division. Damit können nun auch multiplikative Vergleiche angestellt werden: „Heute bist du doppelt so groß wie bei deiner Geburt.“

Mittelwerte und Streuungsmaße wie Varianz und Standardabweichung, die Häufigkeitsverteilungen genauer charakterisieren, können nur aus quantitativen Daten berechnet werden.

Typische Fehlerquellen bei der Datenerhebung

Vor Beginn jeder Erhebung sollten in einem detaillierten Erhebungsplan – unter Berücksichtigung der angepeilten Ziele – die organisatorischen Modalitäten der Erhebung, d. h. der Datengewinnung, festgelegt werden. Fehlerquellen müssen mit den Schülerinnen und Schülern explizit thematisiert werden. So könnte zur Entwicklung und zum Einsatz von Fragebögen beispielsweise mit den Kindern besprochen werden:

- ▶ Verwenden wir ausschließlich geschlossene oder auch offene Fragen? – Erstere ermöglichen eine leichtere Weiterverarbeitung, letztere eignen sich insbesondere, wenn die Antwortkategorien noch nicht bekannt sind und erst aufgrund der erhobenen Daten gefunden werden sollen.
- ▶ Schließen sich die Antwortmöglichkeiten gegenseitig aus oder sind Mehrfachantworten erlaubt? – Letzteres erschwert zwar eindeutig die Datenauswertung, ist aber bei bestimmten Fragen gar nicht anders möglich. Werden Kinder beispielsweise nach



3 Doppelte Anzahl – doppelter Durchmesser – vierfache Fläche! Welche Steigerung beeindruckt mehr?

den Tieren in ihrem Haushalt gefragt, müssen Mehrfachantworten zugelassen werden. Ist nur eine Antwort gewünscht, muss die Frage eher nach dem Lieblingshaustier gestellt werden.

- ▶ Sind die vorgegebenen Antwortmöglichkeiten erschöpfend? – Zwar kann leicht die Kategorie „sonstige“ hinzugefügt werden, diese sollte jedoch bei der Erhebung möglichst selten angekreuzt werden. Wenn von 30 Kindern 18 bei der Frage nach ihrem Lieblingshaustier „sonstige“ angeben, kann man davon ausgehen, dass die Antwortmöglichkeiten nicht angemessen vorgegeben wurden.

▶ Typische Fehlerquellen bei der Datenverarbeitung

Fehler in Strichlisten oder Tabellen erkennen Kinder bei der Auswertung der Fragebögen relativ schnell, wenn die Urliste noch zur Verfügung steht, in der jeder Einzelwert der Befragung tabellarisch (und personenbezogen) aufgeführt ist. Auch „falsche Zahlen“ in Diagrammen werden gefunden.

Subtiler sind „unredliche“ Verzerrungen bzw. die Verwendung des falschen Diagrammtyps. Deshalb lohnt es sich, Diagramme genauer zu betrachten.

In der Grundschule kann mit den Kindern bereits thematisiert werden:

- ▶ Welche Überschrift soll das Diagramm passend zur erhobenen Frage erhalten? Wie sind die Achsen zu beschriften? Welche Fußnoten findet man bei fertigen Diagrammen? Sind Angaben über die Da-

tenquelle und die Anzahl der befragten Personen enthalten?

- ▶ Wie gestaltet man ein Diagramm übersichtlich? Sortiert man die Merkmalsausprägungen oder nicht? – Dies lässt sich mit den Kindern durchaus besprechen, ohne explizit zwischen nominal- und ordinalskalierten Daten zu unterscheiden.
 - ▶ Wie gestaltet man ein Diagramm interessant, ohne es gleichzeitig zu überfrachten? – Kinder können gar nicht früh genug dazu angehalten werden, lieber klare und einfache Diagramme zu verwenden als die durch einen einfachen Knopfdruck zu erzeugenden dreidimensionalen „Spielereien“ ohne größere Aussagekraft.
 - ▶ Wie wirken sich Veränderungen an der Skalierung aus? – Achsen von Diagrammen können gestaucht, gestreckt oder beschnitten werden (s. „Auf einen Blick – alles klar!“ auf S. 32). Auch wenn derartigen Darstellungen kein formaler Fehler nachgewiesen werden kann, so müssen Kinder doch lernen, nicht dem ersten Eindruck zu vertrauen, sondern genau hinzuschauen.
 - ▶ Wie werden Piktogramme angemessen skaliert? – Werden zwei Zahlen dargestellt, von denen eine doppelt so groß wie die andere ist, so muss in einem Säulendiagramm auch die zur zweiten Zahl gehörende Säule doppelt so lang sein. Werden die Zahlen jedoch mittels Flächen veranschaulicht, dann muss der Flächeninhalt verdoppelt werden, d.h. die Seitenlängen bzw. der Durchmesser verändern sich lediglich mit dem Faktor $\sqrt{2}$ **2 3**. Bei einer dreidimensionalen Darstellung verachtfacht sich das Volumen, wenn die Längen verdoppelt werden. Will man also eine doppelt so große Zahl räumlich darstellen, müssen die Längen mit der dritten Wurzel aus zwei multipliziert werden.
- Es kann nicht davon ausgegangen werden, dass die Kinder in einer Unterrichtseinheit all diese Aspekte erfahren und behalten können. Doch wenn Sie kontinuierlich immer wieder Daten und Diagramme unter die Lupe nehmen lassen, werden Sie den Kindern nicht nur einen Zugang zur beschreibenden Statistik eröffnen, sondern sie auf ihrem Weg zu mündigen Bürgerinnen und Bürgern maßgeblich unterstützen.