

Daten erfassen und darstellen

Mathematisches Hintergrundwissen und wichtige Begriffe

11 11

► Daten, Visualisierung, Präsentation

Bereits im Grundschulalter werden Kinder mit vielen Daten konfrontiert, die auf unterschiedliche Art und Weise präsentiert werden. Entsprechend sollten sie die Möglichkeit erhalten, Erfahrungen im Erfassen und Darstellen von Daten zu sammeln, unterschiedliche grafische Darstellungen zu erfassen und zu interpretieren und selbst Informationen in Form grafischer Darstellungen zu verdichten.

VON BERND NEUBERT

Daten (lat. dare = geben) sind logisch gruppierte Informationseinheiten, die auch gespeichert und übertragen werden können. Der Begriff wird in der Regel im Plural verwendet, der Singular „Datum“ (datum = das Gegebene) ist eher ungewöhnlich bzw. hat im Deutschen eine andere Bedeutung (Kalender). Die Verwendung des Plurals weist darauf hin, dass es sich bei Daten immer um eine größere Anzahl handelt. Daten haben einen hohen Informationswert. Die zu vermittelnden Informationen können aber nur dann optimal genutzt werden, wenn sie strukturiert und in geeigneten Darstellungsformen vorliegen.

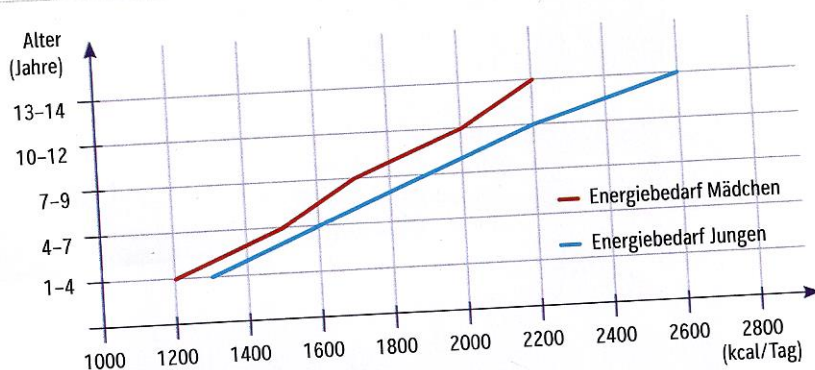
Das Erfassen und Darstellen von Daten erfordert eine Reihe von Überlegungen und Arbeitsschritten. So muss man sich vor dem Erfassen darüber im Klaren sein, was man überhaupt untersuchen möchte und wie dies durchgeführt werden

kann. Nach dem Erfassen der Daten ist zu überlegen, wie die Daten sinnvoll dargestellt werden und welche Schlüsse daraus gezogen werden können.

Das Erfassen von Daten wird auch als statistische Erhebung bezeichnet. Dabei werden die Elemente einer vorgegebenen Menge, die auch als Merkmalsträger oder statistische Einheiten bezeichnet werden, auf eine interessierende Eigenschaft hin betrachtet, die sich aus der Fragestellung ergibt. Die zu untersuchende Menge wird in der beschreibenden Statistik als Grundgesamtheit oder statistische Masse bezeichnet, die Eigenschaft als Identifikationsmerkmal. Für jedes Identifikationsmerkmal müssen mögliche Werte (Kategorien) gefunden werden, die dieses annehmen kann. Diese Werte nennt man Merkmalsausprägungen (Modalitäten).

Für das Beispiel „Geburtsmonat“, das im Anfangsunterricht häufig thematisiert wird, sind die Merkmalsträger die einzelnen Kinder der Klasse. „Geburtsmonat“ ist das Identifikationsmerkmal, die möglichen Werte (Kategorien) sind die Monate Januar bis Dezember, die einzelnen Monate die Merkmalsausprägungen (Modalitäten).

Identifikationsmerkmale können von unterschiedlicher Art sein. Es werden drei Merkmalstypen unterschieden. Qualitative Merkmale (nominal skalierte Merkmale) sind dadurch gekennzeichnet, dass ihre Merkmalsausprägungen nur Beschreibungen sind, d.h., man kann lediglich feststellen, ob diese bei einer statisti-



schen Einheit zutreffen oder nicht. Die Merkmalsausprägungen stehen gleichberechtigt nebeneinander, lassen sich in keine Reihenfolge bringen und sind nicht messbar. Beispiele für qualitative Merkmale sind Haarfarbe, Beruf, Staatsangehörigkeit, Lieblingstier oder Augenzahlen beim Würfeln. Rangmerkmale (ordinal skalierte Merkmale) sind dadurch gekennzeichnet, dass sich ihre Merkmalsausprägungen in eine Reihenfolge bringen lassen. Die Abstände zwischen verschiedenen Ausprägungen sind aber nicht unbedingt gleich und nicht mathematisch interpretierbar. Folglich kann man mit ihnen auch nicht rechnen. Beispiele für Rangmerkmale sind Schulnoten, Güteklassen bei Wein oder Handelsklassen bei Obst. Quantitative Merkmale (metrisch skalierte Merkmale) sind dadurch gekennzeichnet, dass ihre Merkmalsausprägungen in einer festen Reihenfolge angeordnet und die Abstände zwischen den einzelnen Merkmalsausprägungen mathematisch interpretierbar sind. Die quantitativen Merkmale haben als Ausprägungen reelle Zahlen. Erst die quantitativen Merkmale erlauben Vergleiche, Summen- und Differenzenbildungen sowie die Berechnung des arithmetischen Mittels. Beispiele für quantitative Merkmale sind Körpergröße, Alter oder Masse (Gewicht).

Ziel einer statistischen Untersuchung ist, jedem Merkmalsträger aus der Grundgesamtheit eine Merkmalsausprägung des Merkmals zuzuordnen. Um die Auswertung der Daten effektiv zu gestalten, werden häufig vor der Datenerfassung (Planungsphase) mögliche Merkmalsausprägungen vorgegeben. Lässt sich jedem Merkmalsträger eine solche Merkmalsausprägung zuordnen, so heißt ein Merkmal erschöpfend bezüglich der Grundgesamtheit. Bei den Geburtsmonaten ist dies durch die Vorgabe der zwölf Monate eines Jahres problemlos möglich. Bei den Lieblingstieren wird die Vorgabe von Tieren wie Hund, Katze und Pferd als Merkmalsausprägungen in der Regel nicht dazu führen, dass das Merkmal „Lieblingstier“ erschöpfend in der Menge der Kinder einer Klasse wird. Durch Hinzufügen der Merkmalsausprägung „Sonstige“ kann man ein Merkmal stets zu einem erschöpfenden Merkmal machen. Tritt diese Antwort aber zu häufig auf, ist dies Indiz dafür, dass die vorgegebenen Kategorien nicht angemessen ausgewählt wurden.

Vorgehen bei der Erfassung von Daten

Daten können auf unterschiedliche Weise erfasst werden. Wichtige Möglichkeiten der Datenerfassung sind Zählvorgänge (z.B. im Klassenraum oder bei einer Verkehrszählung), Beobachtungen (z.B. Wetterdaten oder Wachstum einer Pflanze) oder Befragungen (mündlich, schriftlich, Telefoninterviews oder internetgestützte Befragungen).

Für Befragungen muss vor allem die Art der Fragen gründlich überlegt werden. So muss man entscheiden, ob geschlossene oder offene Fragen gestellt werden. Bei geschlossenen Fragen werden alle möglichen (oder interessierenden) Antworten ausformuliert und vorgegeben. Die Zahl der Antwortalternativen sollte überschaubar bleiben. Geschlossene Fragen lassen sich aufgrund der vorgegebenen Antwortmöglichkeiten schneller und leichter auswerten als offene Fragen. Eine geschlossene Frage bietet sich für das Beispiel „Geburtsmonat“ an.

Im Gegensatz zu den geschlossenen werden bei offenen Fragen keine Antwortmöglichkeiten vorgegeben. Offene Fragen sind vor allem dann sinnvoll, wenn man wenig über in den Antworten zu erwartende Merkmalsausprägungen voraussagen kann. Eine offene Frage kann beispielsweise nach der Lieblingsspeise gestellt werden.

Offene Fragen erfordern erheblich mehr Aufwand bei der Auswertung der Daten. Zur Auswertung müssen alle Antworten der Befragten in irgendeiner Weise zusammengefasst werden. So ist eine Zusammenfassung zu allgemeineren Merkmalsausprägungen unter Verwendung von Oberbegriffen wie Nudelgerichte, Suppen usw. ratsam.

Auswertung und Darstellung von Daten

Die Auswertung beginnt in der Regel mit der Ermittlung der Häufigkeit für das Auftreten der einzelnen Merkmalsausprägungen. Zählt man für jede Ausprägung aus, wie oft sie auftritt, so erhält man die absolute Häufigkeit dieser Merkmalsausprägung. Für Vergleiche innerhalb einer Datenmenge ist die Angabe der relativen Häufigkeit besser geeignet. Die relative Häufigkeit drückt den Anteil der absoluten Häufigkeit einer Merkmalsausprägung an der Summe der absoluten Häufigkeiten aller Merkmalsausprägungen aus.

Grafische Darstellungen

Um einen besseren Überblick über die erfassten Daten zu bekommen, müssen diese in eine sinnvolle Struktur gebracht werden. Dazu können verschiedene grafische Darstellungsmöglichkeiten genutzt werden. Ziel ist es, wichtige Informationen hervorzuheben. Grafische Darstellungen beschreiben und strukturieren somit das Datenmaterial. Man spricht bei der Beschreibung und Strukturierung der Daten auch von einer Aufbereitung der Daten. Informationen werden so vermittelt, dass der Betrachter Wesentliches möglichst schnell erkennen kann.

Die Aufbereitung von Daten kann mit verschiedenen Zielen verbunden sein. So kann es um einen Gesamtüberblick über die erfassten Daten gehen, in einem anderen Fall interessieren genaue Werte oder ein Vergleich von Einzeler-

gebnissen mit der Gesamtzahl. Schließlich könnte Interesse an der Entwicklung spezieller Daten bestehen. Für verschiedene Zwecke sind einzelne grafische Darstellungsmöglichkeiten unterschiedlich gut geeignet.

Urliste – Strichliste – Tabelle

Die Aufstellung aller ermittelten Daten einer statistischen Erhebung wird als Urliste bezeichnet. Werden die Beobachtungswerte so notiert, wie sie sich nacheinander ergeben, so spricht man von einer „unsortierten Urliste“. Oft werden die Stichprobenwerte aber schon während der Datenerhebung nach möglichen Merkmalsausprägungen geordnet. Dazu verwendet man häufig Strichlisten, die meist in tabellarischer Form angelegt werden.

Eine Tabelle, die jeder Merkmalsausprägung ihre Häufigkeit zuordnet, heißt Häufigkeitstabelle. In ihr werden die absoluten Häufigkeiten bzw. relativen Häufigkeiten der einzelnen Merkmalsausprägungen angegeben. Auch das Betrachten von mehreren Merkmalen und deren Verhältnis zueinander kann in einer Tabelle dargestellt werden. Tabellen geben Informationen in der Regel ohne Datenverlust wieder. Dies ist für Sachverhalte wichtig, bei denen es auf hohe Genauigkeit ankommt und auch Nachkommastellen interessant sind. Tabellen sind aber nicht immer leicht zu lesen. Um einen Überblick über die Verteilung der Daten zu erhalten, muss man eine Tabelle erst genau studieren. Dies stellt einen Nachteil gegenüber anderen grafischen Darstellungsformen, zum Beispiel den verschiedenen Arten an Diagrammen, dar.

Diagramme

Diagramme sind zweidimensionale Darstellungen verschiedener (meist zwei) zusammenhängender Messgrößen. Durch Diagramme werden Zahlen anschaulich gemacht. Nach psychologischen Studien bleiben Informationen, die in einem Diagramm (in Grafiken allgemein) dargestellt sind,

besser im Kurzzeitgedächtnis haften als z.B. Daten, die in Tabellen gezeigt werden. Man sollte bei Diagrammen immer darauf achten, wie man die Daten anordnet. Je deutlicher ein Muster zu erkennen ist, desto besser können Informationen behalten werden.

Der Nachteil von Darstellungen in Diagrammen liegt mitunter in Ungenauigkeiten. Die Werte sind nicht immer ganz genau einzuzeichnen. Dadurch können Teile der Informationen verloren gehen. Deshalb werden bei Diagrammen häufig zusätzlich zur grafischen Darstellung noch die dargestellten Werte in das Diagramm geschrieben.

Stabdiagramm – Streifendiagramm – Balkendiagramm – Säulendiagramm

Stab-, Streifen-, Balken- und Säulendiagramme ermöglichen einen schnellen Vergleich über die dargestellten Daten durch direkte visuelle Wahrnehmung. Diesen Diagrammarten gemeinsam ist die Darstellung in einem rechtwinkligen Koordinatensystem, wobei eine Achse, meist die senkrechte y-Achse, als Skala für die Häufigkeiten dient. Auf der anderen Achse, meist der waagerechten x-Achse, werden die Merkmalsausprägungen notiert.

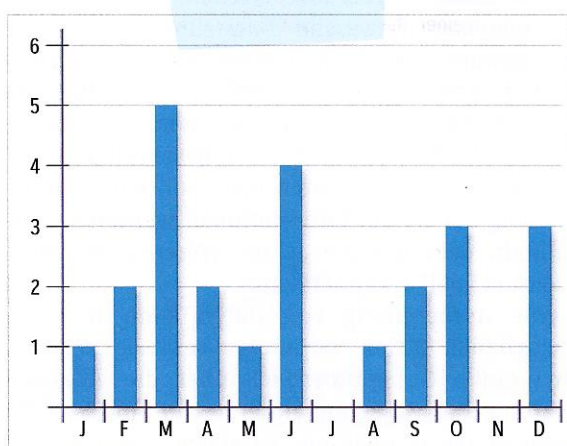
In einem Stabdiagramm geben Streckenlängen (Stablängen) die absolute bzw. relative Häufigkeit der Merkmalsausprägung an. Stabdiagramme sind sehr anschaulich und ermöglichen einen schnellen Überblick über die Verteilung der erfassten Daten. Vor allem Vergleiche zwischen Häufigkeiten der dargestellten Merkmalsausprägungen sind schnell möglich. Bei qualitativen Merkmalen kann die Einteilung der waagerechten Achse (Merkmalsausprägungen) beliebig gewählt werden. Aus optischen Gründen ist es allerdings zu empfehlen, gleich große Abstände zwischen den Stäben zu lassen.

Bei Rangmerkmalen hat die Einteilung der Anordnung der Merkmalsausprägungen zu entsprechen. Zur „optischen Aufbesserung“ werden häufig Rechtecke bzw. Quader als Stäbe verwendet. Es entstehen Streifen- bzw. Säulendiagramme. Bei Streifendiagrammen – sie werden auch als Balkendiagramme bezeichnet – wird für die Häufigkeit jeder Merkmalsausprägung statt einer Strecke ein Rechteck eingezeichnet. Alle Rechtecke sollten dieselbe Breite haben, da man vordergründig nicht die Höhe der Rechtecke, sondern die Fläche wahrnimmt.

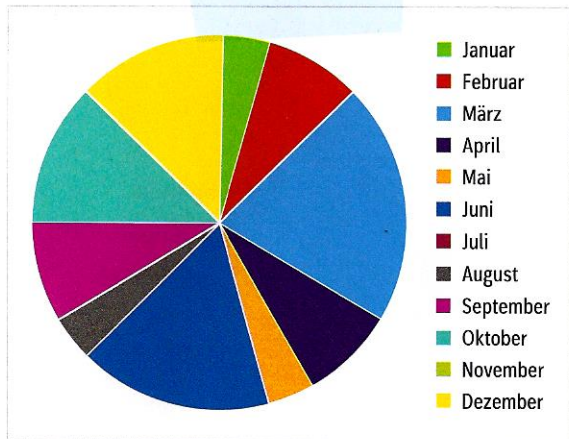
Abbildung 1 zeigt ein Streifendiagramm mit fiktiven Werten für das Beispiel „Geburtsmonat“.

Kreis- und Blockdiagramm

Kreis- und Blockdiagramme nutzen die Fläche als Mittel der Veranschaulichung. Es erfolgt eine Zuordnung jeder Merkmalsausprägung zu einem Kreissektor bzw. Teilrechteck.



1 Ein Streifendiagramm zum Beispiel „Geburtsmonat“



2 Kreisdiagramm zum Beispiel „Geburtsmonat“

Bei Kreisdiagrammen (siehe Abb. 2) wird die Fläche eines Kreises in Kreissektoren unterteilt, wobei jeder Kreissektor eine Merkmalsausprägung veranschaulicht. Da die Kreisflächen sich proportional zu den Winkeln der Kreissektoren verhalten, sind diese die wesentlichen Parameter zum Zeichnen eines Kreisdiagramms. Die Größe des Winkels für einen Kreissektor lässt sich als Produkt aus der relativen Häufigkeit der Merkmalsausprägung und dem Vollwinkel des Kreises (360°) berechnen. Bei Blockdiagrammen werden anstelle der Kreissektoren Rechtecke verwendet.

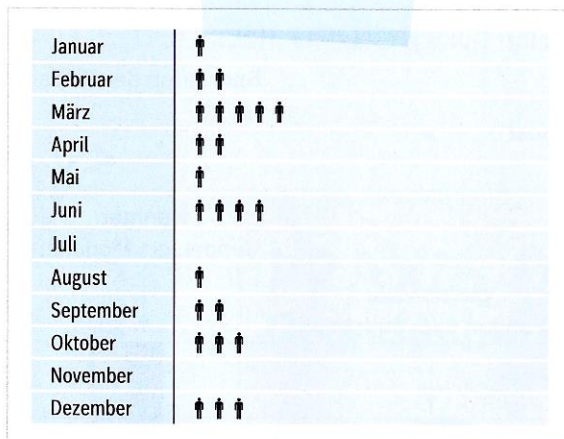
Aus Kreis- und Blockdiagrammen lassen sich Beziehungen zwischen der Gesamtanzahl der Ergebnisse (gesamte Kreis- bzw. Rechteckfläche) und den jeweiligen Einzelergebnissen (Kreissektoren bzw. Teilrechtecke) schnell erfassen. Auch die relativen Häufigkeiten der Merkmalsausprägungen kann man vergleichen, absolute Häufigkeiten werden dagegen nicht ersichtlich.

In Abbildung 2 sind die Werte aus dem Streifendiagramm zum Thema „Geburtsmonat“ in einem Kreisdiagramm dargestellt.

Piktogramm

Unter einem Piktogramm – in Grundschulbüchern häufig als Bilddiagramm bezeichnet – versteht man die Veranschaulichung von Häufigkeiten anhand von Symbolen, Figuren oder Abbildungen zur Darstellung von statistischen Daten (siehe Abb. 3). Bei kleinen Anzahlen repräsentiert ein Symbol in der Regel eine grundlegende Einheit der Merkmalsausprägung. Die Bilder stellen damit einen direkten Bezug zu den dargestellten Zahlen her.

Piktogramme werden aber auch verwendet, wenn größere Anzahlen als im Beispiel veranschaulicht werden sollen. Für das Übertragen in ein Piktogramm werden die Werte gerundet und für eine größere (gerundete) Anzahl wird ein Symbol gewählt; mitunter wird auch mit „anteiligen Symbolen“ gearbeitet.

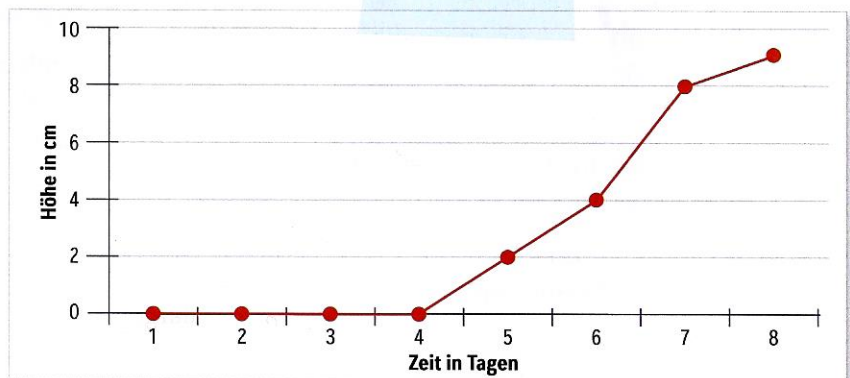


3 Piktogramm zum Beispiel „Geburtsmonat“

Linien- bzw. Kurvendiagramm

Die Darstellung von zeitlichen Abläufen ist in Linien- bzw. Kurvendiagrammen gut möglich. In ein rechtwinkliges Koordinatensystem werden auf der Ordinate (x-Achse) die Zeit und als Abszissenwert (y-Achse) die zugehörige Merkmalsausprägung (häufig ein Messwert) eingetragen.

Handelt es sich um metrische Merkmale, so können die einzelnen Punkte zu einer Kurve verbunden werden. In unserem Beispiel handelt es sich zwar um isolierte Messwerte, die jedoch verbunden werden können: Beim Wachstum der Pflanze kann davon ausgegangen werden, dass es sich gleichmäßig vollzieht.



4 Wachstum einer Pflanze

In Liniendiagrammen werden Trends von Entwicklungen schnell sichtbar. Zu Vergleichszwecken können auch mehrere Datenmengen in dasselbe Koordinatensystem gezeichnet werden, indem man z.B. verschiedene Kurven in verschiedenen Farben einzeichnet. ■

Der Autor

Dr. Bernd Neubert ist Oberstudienrat im Hochschuldienst am Institut für Didaktik der Mathematik der Justus Liebig Universität Gießen.