Lernzettel - Statistik und Daten

Klasse: Klasse 8

In der Statistik untersuchst du Daten wie Noten, Körpergrößen oder Umfrageergebnisse. Du ordnest sie, stellst sie dar und beschreibst sie mit Kenngrößen.

Einstieg: Die Mathe-Arbeit

Die Klasse 8a hat die Mathe-Arbeit zurück. Zwei Kinder bekamen eine 1, sechs eine 2, acht eine 3, fünf eine 4, zwei eine 5 und eines eine 6. Die Lehrerin sagt: „Der Durchschnitt liegt bei $3{,}1$.“ Jonas meint: „Aber die meisten haben eine 3.“ Und Mia: „Die Hälfte der Klasse hat 3 oder besser.“ Alle drei haben recht, denn sie beschreiben dieselben Daten mit verschiedenen Kenngrößen: Mittelwert, Modus und Median. In der Statistik lernst du, Daten zu ordnen, mit Diagrammen darzustellen und mit Kenngrößen zu beschreiben. Dabei merkst du: Welche Kenngröße gut passt, hängt von den Daten ab, und eine Grafik kann ehrlich oder täuschend sein.

Daten ordnen

Definition: Urliste und Häufigkeiten

Die einzelnen Werte heißen **Urliste**. Die **absolute Häufigkeit** gibt an, wie oft ein Wert vorkommt. Die **relative Häufigkeit** ist der Anteil an allen Daten: $h = \frac{\text{absolute Häufigkeit}}{\text{Anzahl aller Daten}}$. Du gibst sie als Bruch, Dezimalzahl oder in Prozent an.

Beispiel 1: Umfrage

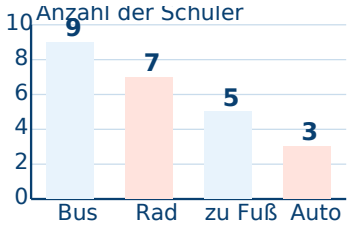
12 von 30 Schülern wählen Fußball. Wie groß ist die relative Häufigkeit?

$\frac{12}{30} = 0{,}4 = 40\,\%$.

Diagramme

Regel: Diagrammtypen

- **Säulendiagramm** (Balkendiagramm): gut für den Vergleich von Häufigkeiten.
- **Kreisdiagramm**: zeigt Anteile am Ganzen. Der Winkel eines Ausschnitts ist $h \cdot 360^\circ$; 40 % entsprechen 144°.
- **Liniendiagramm**: gut für Entwicklungen über die Zeit.



\n

Achtung

Achte beim Lesen von Diagrammen auf die Achsenbeschriftung. Eine abgeschnittene y-Achse lässt Unterschiede größer erscheinen, als sie sind. Bei Kreisdiagrammen muss die Winkelsumme 360° ergeben.

Kenngrößen der Lage

Definition: Mittelwert, Median, Modus

- Das **arithmetische Mittel** ist die Summe aller Werte geteilt durch ihre Anzahl: $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$.
- Der **Median** (Zentralwert) ist der mittlere Wert der geordneten Daten. Bei gerader Anzahl nimmst du den Mittelwert der beiden mittleren Werte.
- Der **Modus** ist der häufigste Wert.

Beispiel 2: Kenngrößen

Bestimme Mittelwert der Daten 4, 7, 5, 9, 5 sowie Median und Modus von 11, 12, 14, 15, 15, 18.

Mittelwert: Summe 30, $30 : 5 = 6$.

Median: $\frac{14+15}{2} = 14{,}5$. Modus: 15.

Streuung

Definition: Spannweite

Die **Spannweite** ist der Unterschied zwischen größtem und kleinstem Wert: $R = x_{\max} - x_{\min}$. Sie zeigt, wie weit die Daten auseinanderliegen.

Beispiel 3: Gleiches Mittel, andere Streuung

Klasse A hat die Noten 2, 3, 3, 3, 3, 4, Klasse B die Noten 1, 1, 3, 3, 5, 5. Vergleiche.

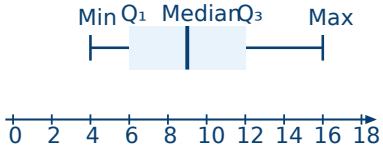
Beide Klassen haben das Mittel 3.

Die Spannweite ist 2 bei Klasse A und 4 bei Klasse B. Klasse B streut stärker.

Quartile und Boxplot

Definition: Quartile

Der Median teilt die geordneten Daten in zwei Hälften. Der Median der unteren Hälfte heißt **unteres Quartil** Q_1, der Median der oberen Hälfte **oberes Quartil** Q_3. Der **Quartilsabstand** $Q_3 - Q_1$ enthält die mittleren 50 % der Daten.



\n

So gehst du vor: Boxplot zeichnen

1. Ordne die Daten der Größe nach.
2. Bestimme Minimum, Maximum und Median.
3. Bestimme Q_1 und Q_3 als Mediane der unteren und der oberen Hälfte.
4. Zeichne die Box von Q_1 bis Q_3 mit einem Strich beim Median und die Linien (Whisker) bis Minimum und Maximum.

Beispiel 4: Quartile

Daten: 2, 3, 5, 6, 8, 9, 11, 14. Bestimme die Fünf-Zahlen-Zusammenfassung und den Quartilsabstand.

Median: $\frac{6+8}{2} = 7$.

Untere Hälfte 2, 3, 5, 6: $Q_1 = 4$. Obere Hälfte 8, 9, 11, 14: $Q_3 = 10$. Minimum 2, Maximum 14.

Quartilsabstand: $10 - 4 = 6$.

Ausreißer und gewichtetes Mittel

Merke: Ausreißer

Ein sehr großer oder kleiner Wert, ein **Ausreißer**, verändert das arithmetische Mittel stark, den Median kaum.

Beispiel 5: Taschengeld

Das Taschengeld beträgt 10, 12, 15, 11 und 100 €. Welche Kenngröße beschreibt die Gruppe besser?

Mittelwert: $\frac{148}{5} = 29{,}60$ €. Median: 12 €. Hier beschreibt der Median die Gruppe besser.

Beispiel 6: Gewichtetes Mittel

Zwei Klassenarbeiten (Note 2 und 3) zählen doppelt, ein Test (Note 1) einfach. Wie lautet der Schnitt?

$$\frac{2 \cdot 2 + 2 \cdot 3 + 1}{5} = \frac{11}{5} = 2{,}2.$$

Daten sinnvoll interpretieren

Hinweis

Frage dich: Wer wurde befragt? Wie viele Personen sind es? Wurde zufällig ausgewählt? Eine Umfrage nur unter Fußballfans sagt wenig über die ganze Schule. Für Einkommen oder Taschengeld nimmt man oft den Median, für Noten häufig den Mittelwert, für „Welche Schuhgröße wird am meisten verkauft?“ den Modus.

Mittelwert aus einer Häufigkeitstabelle

Wenn viele gleiche Werte vorkommen, ist es mühsam, alle Werte aufzuschreiben. Du rechnest dann mit der Häufigkeitstabelle: Jeden Wert mit seiner Häufigkeit multiplizieren, die Produkte addieren und durch die Gesamtzahl der Daten teilen. Das ist ein gewichtetes Mittel, bei dem die Häufigkeiten die Gewichte sind.

Beispiel 7: Mittelwert aus der Tabelle

In einer Klasse wurden die Geschwisterzahlen erfragt: 0 Geschwister bei 5 Kindern, 1 bei 11, 2 bei 6, 3 bei 2. Berechne den Mittelwert.

Gesamtzahl: $5 + 11 + 6 + 2 = 24$ Kinder. Summe: $0 \cdot 5 + 1 \cdot 11 + 2 \cdot 6 + 3 \cdot 2 = 29$. Mittelwert: $\frac{29}{24} \approx 1{,}21$ Geschwister pro Kind. Der Median liegt bei 1, denn der $12.$ und der $13.$ Wert sind 1.

Boxplots lesen und vergleichen

Ein Boxplot zeigt auf einen Blick, wie die Daten verteilt sind. Die Box enthält die mittleren $50\,\%$ der Daten, der Strich darin ist der Median, und die Linien (Whisker) reichen bis zum Minimum und Maximum. Ist die Box schmal, liegen viele Daten eng beieinander. Liegt der Median nicht in der Mitte der Box, sind die Daten schief verteilt. Beim Vergleich zweier

Boxplots schaust du auf Lage (Median), Streuung (Länge der Box und Whisker) und Ausreißer.

Beispiel 8: Zwei Boxplots vergleichen

Zwei Läufer haben diese Fünf-Zahlen-Zusammenfassungen (Zeiten in s): Läufer A: $12; 13; 14; 15; 17$. Läufer B: $11; 13; 15; 17; 21$. Wer ist gleichmäßiger?

Für A gilt $Q_3 - Q_1 = 15 - 13 = 2$, für B $17 - 13 = 4$. A hat die kleinere Box, ist also gleichmäßiger. B hat aber die schnellste Zeit (11 s) und die langsamste (21 s).

Statistiken kritisch lesen

Zahlen können täuschen, auch wenn sie richtig berechnet sind. Prüfe immer: Wie viele Personen wurden befragt? Wurden sie zufällig ausgewählt? Beginnt die Achse bei null? Wird der Mittelwert angegeben, obwohl es Ausreißer gibt? Eine abgeschnittene y-Achse lässt kleine Unterschiede riesig wirken. Eine Umfrage nur unter Fußballfans sagt wenig über die ganze Schule. Je größer und zufälliger die Stichprobe, desto verlässlicher das Ergebnis.

Beispiel 9: Stichprobe beurteilen

Ein Hersteller wirbt: „9 von 10 Kunden sind zufrieden.“ Befragt wurden 10 Personen aus dem Freundeskreis der Geschäftsleitung. Wie beurteilst du das?

Die Stichprobe ist sehr klein ($n = 10$) und nicht zufällig. Der Anteil $90\,\%$ sagt wenig aus. Besser wären mehrere hundert zufällig ausgewählte Kunden.

Achtung: Typische Fehler

Erstens: Den Median bei gerader Anzahl als einen der beiden mittleren Werte abzulesen statt ihren Mittelwert zu bilden. Zweitens: Die Daten nicht zu ordnen. Drittens: Bei den Quartilen den Median bei ungerader Anzahl in beide Hälften zu nehmen. Viertens: Mittelwert und Median zu verwechseln.

Wusstest du?

Der Boxplot wurde 1970 vom amerikanischen Statistiker John Tukey erfunden. Er nannte ihn zunächst „Box-and-Whisker-Plot“. Statistik begegnet dir überall: bei Wahlprognosen, in der Medizin und bei Streaming-Diensten, die dir Filme vorschlagen, weil sie die Daten vieler Nutzer vergleichen.

Merke: Das Wichtigste auf einen Blick

- Absolute Häufigkeit: wie oft ein Wert vorkommt; relative Häufigkeit: Anteil an allen Daten (Bruch, Dezimalzahl oder Prozent).
- Mittelwert: Summe geteilt durch Anzahl; Median: mittlerer Wert der geordneten

Daten; Modus: häufigster Wert.
- Spannweite $R = x_{\max} - x_{\min}$ zeigt, wie weit die Daten auseinanderliegen.
- Quartile Q_1 und Q_3 teilen die geordneten Daten in Viertel; der Quartilsabstand $Q_3 - Q_1$ enthält die mittleren $50\,\%$.
- Boxplot: Minimum, Q_1, Median, Q_3, Maximum.
- Ausreißer verändern den Mittelwert stark, den Median kaum.
- Gewichtetes Mittel: Werte mit ihren Gewichten multiplizieren, summieren und durch die Gewichtssumme teilen.
- Kreisdiagramm: Winkel $= h \cdot 360^\circ$; Diagramme kritisch lesen (Achsen, Stichprobe).

Achtung

Ordne die Daten, bevor du den Median bestimmst. Bei gerader Anzahl nimmst du zwei mittlere Werte. Verwechsle Mittelwert, Median und Modus nicht. Die Spannweite ist eine Differenz, keine Summe.

Vertiefung: Daten kritisch auswerten

So gehst du vor: Eine Datenreihe vollständig auswerten

1. Ordne die Daten der Größe nach und zähle, wie viele Werte es sind.
2. Bestimme die Lagemaße: Mittelwert, Median und Modus.
3. Bestimme die Streumaße: Spannweite, Quartile und Quartilsabstand.
4. Suche nach Ausreißern, also Werten, die weit von den übrigen entfernt liegen.
5. Wähle ein passendes Diagramm: Säulen für Häufigkeiten, Kreis für Anteile, Boxplot für den Vergleich von Gruppen.
6. Formuliere eine Aussage in ganzen Sätzen und nenne dabei auch die Grenzen deiner Daten, etwa eine kleine oder nicht zufällige Stichprobe.

Beispiel: Gleicher Mittelwert, verschiedene Streuung

Zwei Wetterstationen messen eine Woche lang die Mittagstemperatur. Station A: 20, 21, 22, 22, 23, 24, 22. Station B: 12, 28, 15, 30, 18, 26, 25. Vergleiche.

Beide Stationen haben die Summe 154 und damit den Mittelwert 22 °C. Die Spannweite ist aber bei A nur 4 °C, bei B dagegen 18 °C. Wer nur den Mittelwert betrachtet, hält beide Orte für gleich. Erst die Streuung zeigt, dass das Wetter bei B viel wechselhafter ist. Deshalb gibt man zu einem Lagemaß immer auch ein Streumaß an.

Ottos Tipp

Ordne die Daten immer zuerst. Dann siehst du sofort Minimum, Maximum, Median und

Modus. Ein Mittelwert muss immer zwischen dem kleinsten und dem größten Wert liegen. Frage dich bei jedem Ergebnis: Passt der Wert zu den Daten?