Lernzettel - Statistik: Mittelwert, Median und Boxplot

Klasse: Klasse 7

Statistiken begegnen dir bei Noten, Umfragen und im Sport. Hier lernst du Lage- und Streuungsmaße kennen und zeichnest Boxplots.

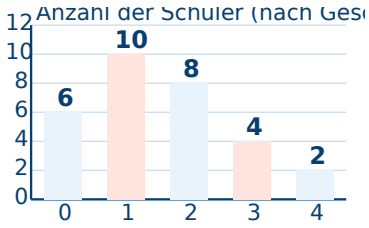
Einstieg: Die Schulweg-Umfrage

Die Klasse 7a hat gemessen, wie viele Minuten jedes Kind zur Schule braucht: 5, 10, 10, 15, 15, 15, 20, 45. Wie lang ist der Schulweg „typischerweise“? Der Mittelwert ist $\frac{135}{8} \approx 16{,}9$ Minuten, der Median 15 Minuten, der häufigste Wert ebenfalls 15. Der Ausreißer 45 (jemand wohnt weit weg) zieht den Mittelwert nach oben. Je nachdem, welches Maß du wählst, erzählt dieselbe Tabelle eine etwas andere Geschichte.

Daten sammeln und darstellen

Definition: Urliste und Häufigkeiten

In der **Statistik** werden Daten erhoben, geordnet und ausgewertet. Die Rohdaten stehen in einer **Urliste**, zum Beispiel die Noten einer Klassenarbeit: $2; 3; 1; 4; 2; 3; 3; 5$. Die **absolute Häufigkeit** zählt, wie oft ein Wert vorkommt, die **relative Häufigkeit** gibt den Anteil an der Gesamtzahl an.



Beispiel 1: Häufigkeitstabelle

In einer Klasse mit 30 Schülern haben 6 kein Geschwisterkind, 10 eines, 8 zwei, 4 drei und 2 vier. Bestimme Mittelwert und Median.

Mittelwert: $\frac{0\cdot 6+1\cdot 10+2\cdot 8+3\cdot 4+4\cdot 2}{30} = \frac{46}{30} \approx 1{,}53$. Die 15. und 16. Zahl sind beide 1, also ist der Median 1.

Lagemaße: Mittelwert, Median, Modus

Definition: Mittelwert

Der **Mittelwert** (arithmetisches Mittel) ist die Summe aller Werte geteilt durch ihre

Anzahl: $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$.

Definition: Median

Der **Median** ist der mittlere Wert der *der Größe nach geordneten* Liste. Bei gerader Anzahl nimmt man den Mittelwert der beiden mittleren Werte.

Definition: Modus

Der **Modus** (Modalwert) ist der am häufigsten vorkommende Wert.

Beispiel 2: Noten

Noten 1, 2, 2, 3, 3, 3, 4, 5 (geordnet). Bestimme Mittelwert, Median und Modus.

Mittelwert: $\frac{23}{8} = 2{,}875$. Median: Die beiden mittleren Werte sind 3 und 3, also 3. Modus: 3.

Beispiel 3: Ausreißer

Monatsgehälter 1 800, 2 000, 2 100, 2 200 und 9 000 Euro. Welches Lagemaß passt besser?

Mittelwert: $\frac{17\,100}{5} = 3\,420$. Median: 2 100. Der Ausreißer 9 000 zieht den Mittelwert nach oben, der Median bleibt robust. Hier ist der Median aussagekräftiger.

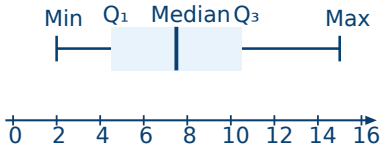
Streuungsmaße: Spannweite und Quartile

Definition: Spannweite

Die **Spannweite** ist der Unterschied zwischen größtem und kleinstem Wert: $R = x_{\max} - x_{\min}$.

Definition: Quartile

Der Median teilt die geordneten Daten in zwei Hälften. Der **untere Quartil** Q_1 ist der Median der unteren Hälfte, der **obere Quartil** Q_3 der Median der oberen Hälfte. Der **Quartilsabstand** $Q_3 - Q_1$ beschreibt die Streuung der mittleren $50\,\%$ der Daten.



So gehst du vor: Quartile bestimmen

1. Ordne die Daten der Größe nach.
2. Bestimme den Median.
3. Bestimme Q_1 als Median der unteren Hälfte und Q_3 als Median der oberen Hälfte. Bei ungerader Anzahl gehört der Median nicht dazu.
4. Berechne $Q_3 - Q_1$.

Beispiel 4: Quartile

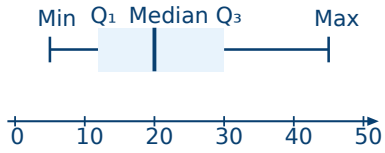
Daten $5, 8, 10, 10, 12, 15, 20, 25, 30$ ($n = 9$). Bestimme Median, Quartile, Quartilsabstand und Spannweite.

Median 12 (der fünfte Wert). Untere Hälfte $5, 8, 10, 10$: $Q_1 = \frac{8+10}{2} = 9$. Obere Hälfte $15, 20, 25, 30$: $Q_3 = \frac{20+25}{2} = 22{,}5$. Quartilsabstand 13,5. Spannweite 25.

Der Boxplot

Definition: Boxplot

Ein **Boxplot** (Kastenschaubild) zeigt fünf Werte: Minimum, Q_1, Median, Q_3, Maximum. Die Box reicht von Q_1 bis Q_3 und enthält den Median als Strich. Die Linien (Whisker) führen zum Minimum und zum Maximum. In der Box liegen die mittleren $50\,\%$ der Daten.



Beispiel 5: Fünf-Punkte-Zusammenfassung

Daten $2, 4, 5, 7, 8, 9, 12, 15$ ($n = 8$).

Median $\frac{7+8}{2} = 7{,}5$. $Q_1 = \frac{4+5}{2} = 4{,}5$, $Q_3 = \frac{9+12}{2} = 10{,}5$. Zusammenfassung: $2 \mid 4{,}5 \mid 7{,}5 \mid 10{,}5 \mid 15$.

Hinweis

Mit Boxplots lassen sich Datensätze gut vergleichen.

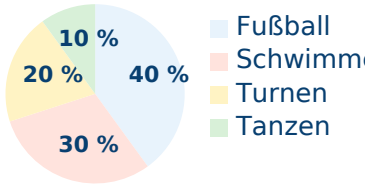
Daten auswerten und Diagramme lesen

Merke: Lage und Streuung

Zwei Datensätze können denselben Mittelwert haben und sich trotzdem unterscheiden: $2, 3, 3, 3, 4$ und $1, 1, 3, 5, 5$ haben beide den Mittelwert 3, aber die Spannweiten 2 und 4. Beschreibe deshalb immer Lage *und* Streuung.

Regel: Diagrammarten

Ein **Säulendiagramm** eignet sich für den Vergleich einzelner Werte, ein **Kreisdiagramm** für Anteile, ein **Liniendiagramm** für Entwicklungen über die Zeit. Beim Kreisdiagramm gilt: $\text{Winkel} = \frac{\text{Anteil}}{100\,\%} \cdot 360^\circ$. Also bedeuten $25\,\%$ einen Winkel von 90°.



Hinweis

Beginne in Säulendiagrammen die Achsen bei null, sonst wirken kleine Unterschiede groß. Beschrifte beide Achsen mit Größe und Einheit und gib dem Diagramm eine Überschrift. Der Notendurchschnitt ist ein Mittelwert. Frage bei Statistiken immer: Wie viele wurden befragt? Mittelwert oder Median? Gibt es Ausreißer?

Beispiel 6: Im Alltag (Handyakku-Test)

Ein Magazin testet die Laufzeit von elf Handys (in Stunden): $9; 11; 12; 12; 13; 14; 14; 15; 16; 18; 22$. Beschreibe die Daten.

Median 14 (sechster Wert). $Q_1 = 12$, $Q_3 = 16$. Quartilsabstand 4, Spannweite 13. Mittelwert $\frac{156}{11} \approx 14{,}2$. Die Hälfte der Handys hält länger als 14 Stunden, die mittleren $50\,\%$ liegen zwischen 12 und 16 Stunden.

Beispiel 7: Aus der Häufigkeitstabelle

Eine Klasse (20 Schüler) hat folgende Geschwisterzahlen: 0: 4 Kinder, 1: 8, 2: 5, 3: 3. Berechne Mittelwert und Median.

Mittelwert $\frac{0\cdot 4 + 1\cdot 8 + 2\cdot 5 + 3\cdot 3}{20} = \frac{27}{20} = 1{,}35$. Median: Der $10.$ und $11.$ Wert sind beide 1, also Median 1. Modus 1.

Beispiel 8: Ein Boxplot wird gezeichnet

Zeichne den Boxplot für $2; 4; 5; 7; 8; 9; 12; 15$ (Fünf-Punkte-Zusammenfassung $2 \mid 4{,}5 \mid 7{,}5 \mid 10{,}5 \mid 15$).

Lege eine Zahlengerade von 0 bis 16 an. Zeichne die Box von $4{,}5$ bis $10{,}5$, den Median-

Strich bei $7{,}5$ und die Whisker bis 2 und 15. Die Box ist 6 Einheiten breit, die Gesamtbreite 13.

Typische Fehler

(1) Den Median bestimmen, ohne die Daten zu ordnen. (2) Bei gerader Anzahl nur einen mittleren Wert nehmen. (3) Bei ungerader Anzahl den Median in die Hälften einrechnen, wenn man Quartile bestimmt. (4) Mittelwert, Median und Modus verwechseln. (5) Säulendiagramme mit abgeschnittener Achse zeichnen.

Wusstest du?

Den Boxplot erfand der amerikanische Statistiker John W. Tukey im Jahr 1970. Das Wort „Statistik“ kommt vom lateinischen „status“, dem Staat: Früher zählten Regierungen Einwohner und Ernten. Eine bekannte Stolperfalle ist der „Durchschnittsmensch“: Der Mittelwert passt oft zu niemandem wirklich.

Merke: Das Wichtigste auf einen Blick

- Mittelwert: Summe geteilt durch Anzahl; Median: Mitte der geordneten Liste; Modus: häufigster Wert.
- Der Median ist robust gegen Ausreißer.
- Spannweite $= x_{\max} - x_{\min}$; Quartilsabstand $= Q_3 - Q_1$.
- Boxplot: Minimum, Q_1, Median, Q_3, Maximum.
- In der Box liegen die mittleren $50\,\%$ der Daten.
- Datensätze mit gleichem Mittelwert können unterschiedlich streuen.

Achtung

Der Median darf nur an geordneten Daten bestimmt werden. Beim Mittelwert durch die richtige Anzahl teilen. Bei gerader Anzahl nimmt man die beiden mittleren Werte. Quartile und halbe Spannweite sind nicht dasselbe.

Vertiefung: Daten richtig auswerten

Einstieg: Das durchschnittliche Gehalt

In einer kleinen Firma arbeiten neun Angestellte, die jeweils *2500\,€* im Monat verdienen, und die Chefin, die *25\,000\,€* verdient. Das durchschnittliche Gehalt beträgt *(9\cdot2500+25\,000):10=4750\,€*. Würde die Firma damit werben, „bei uns verdient man im Schnitt 4750 €“, wäre das zwar richtig gerechnet, aber irreführend, denn neun von zehn Personen verdienen viel weniger. Der Median beträgt dagegen *2500\,€* und beschreibt das typische Gehalt viel besser. Dieses Beispiel zeigt: Wer Daten auswertet, muss entscheiden, welches Lagemaß zur Frage passt.

Merke: Welches Lagemaß wann?

Der **Mittelwert** berücksichtigt jeden einzelnen Wert. Er eignet sich, wenn die Daten gleichmäßig verteilt sind und keine Ausreißer enthalten, zum Beispiel bei Temperaturen. Der **Median** ist unempfindlich gegen Ausreißer und eignet sich bei schiefen Verteilungen wie Einkommen oder Wartezeiten. Der **Modus** ist das einzige Lagemaß, das man auch bei Merkmalen ohne Reihenfolge verwenden kann, etwa bei Lieblingsfarben oder Sportarten. Gib bei einer Auswertung am besten mehrere Kennwerte an.

So gehst du vor: Einen Boxplot zeichnen

1. Ordne die Daten der Größe nach und bestimme Minimum, Q_1, Median, Q_3 und Maximum.
2. Zeichne eine Zahlengerade mit passender Einteilung, die vom Minimum bis zum Maximum reicht.
3. Zeichne eine Box von Q_1 bis Q_3 und darin einen senkrechten Strich beim Median.
4. Verbinde die Box mit Linien („Antennen“ oder „Whisker“) bis zum Minimum und bis zum Maximum.
5. Beschrifte die Achse mit Einheit. In der Box liegt die mittlere Hälfte der Daten, links und rechts davon jeweils ein Viertel.

Beispiel 9: Fehlender Wert

Vier Tests ergaben 12, 15, 9 und 14 Punkte. Wie viele Punkte braucht man im fünften Test für einen Durchschnitt von 13 Punkten?

Ein Durchschnitt von 13 bei fünf Tests bedeutet eine Gesamtsumme von $5 \cdot 13 = 65$ Punkten. Bisher sind es $12 + 15 + 9 + 14 = 50$ Punkte. Im fünften Test sind also $65 - 50 = 15$ Punkte nötig. Merke: Bei Aufgaben mit gesuchtem Wert rechnest du immer über die Summe, denn Mittelwert mal Anzahl ergibt die Summe aller Werte.

Achtung: Median nur bei geordneten Daten

Der häufigste Fehler beim Median ist, den mittleren Wert der ungeordneten Liste zu nehmen. Bei $3;\ 9;\ 1;\ 7;\ 5$ ist der mittlere Eintrag die 1, der Median ist aber 5. Ordne die Daten deshalb immer zuerst. Achte außerdem darauf, dass bei einer geraden Anzahl der Median der Mittelwert der beiden mittleren Werte ist.

Schon gewusst?

Der Boxplot wurde 1970 von dem amerikanischen Statistiker John Tukey erfunden. Er wollte ein Diagramm, das man schnell von Hand zeichnen kann und das trotzdem viel über die Daten verrät: Lage, Streuung und Ausreißer auf einen Blick. Heute findet man Boxplots in Forschungsberichten, Medizinstudien und Sportanalysen. Besonders

praktisch sind sie, um mehrere Gruppen nebeneinander zu vergleichen.

Beispiel: Zwei Boxplots vergleichen

Die Sprungweiten zweier Gruppen im Weitsprung werden als Boxplots dargestellt. Gruppe A: Median 3,6 m, Box von 3,3 m bis 3,9 m. Gruppe B: Median 3,8 m, Box von 3,0 m bis 4,4 m. Vergleiche.

Gruppe B springt im Mittel weiter, denn ihr Median liegt 20 cm höher. Ihre Box ist aber doppelt so breit, die Leistungen streuen also viel stärker. In Gruppe A sind die Kinder ähnlich gut, in Gruppe B gibt es sehr starke und eher schwache Springer. Beim Vergleich von Boxplots betrachtet man deshalb immer Lage und Breite gemeinsam.

Ottos Tipp

Ordne die Daten immer zuerst der Größe nach und nummeriere sie durch. Dann findest du Median und Quartile ohne Fehler. Prüfe den Mittelwert: Er muss zwischen dem kleinsten und dem größten Wert liegen.