

Deskriptive Kennwerte

statistik-online.ch

Deskriptive Kennwerte fassen Daten kurz und verständlich zusammen. Sie beschreiben, wo die Werte ungefähr liegen, wie stark sie streuen und ob einzelne Werte besonders auffällig sind.

1 Grundidee

Deskriptive Statistik beschreibt vorhandene Daten. Sie testet noch keine Hypothesen und zieht noch keine Schlüsse auf eine Population.

Typische Fragen sind:

- Wo liegt ein typischer Wert?
- Wie stark unterscheiden sich die Werte voneinander?
- Gibt es auffällig kleine oder grosse Werte?
- Welche Kennwerte passen zur Fragestellung?

2 Lagekennwerte

Lagekennwerte beschreiben, wo die Daten ungefähr liegen.

Arithmetisches Mittel

Das arithmetische Mittel ist der Durchschnitt aller Werte.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n}$$

Wichtig: Das arithmetische Mittel reagiert stark auf Ausreisser.

Median

Der Median ist der mittlere Wert, nachdem die Daten sortiert wurden. Es liegen 50% der Werte darunter und 50% der Werte darüber.

- Bei ungerader Anzahl Werte: der Wert in der Mitte.
- Bei gerader Anzahl Werte: Durchschnitt der zwei mittleren Werte.

$$\text{Median} = \begin{cases} x_{\frac{n+1}{2}}, & \text{wenn } n \text{ ungerade ist} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{wenn } n \text{ gerade ist} \end{cases}$$

Wichtig: Der Median ist robust gegen Ausreisser.

Modus

Der Modus ist der Wert, der am häufigsten vorkommt. Es kann keinen, einen oder mehrere Modi geben.

3 Streuungskennwerte

Streuungskennwerte beschreiben, wie weit die Werte auseinanderliegen.

Minimum, Maximum und Spannweite

Das Minimum ist der kleinste Wert im Datensatz. Das Maximum ist der grösste Wert im Datensatz.

$$\text{Spannweite} = \text{Maximum} - \text{Minimum}$$

Die Spannweite ist einfach zu berechnen, reagiert aber stark auf extreme Werte.

Varianz und Standardabweichung

Die Varianz beschreibt die durchschnittliche quadrierte Abweichung vom Mittelwert. Die Standardabweichung ist die Wurzel aus der Varianz. Inhaltlich beschreibt sie die durchschnittliche Abweichung der Werte vom Mittelwert.

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

$$s = \sqrt{s^2}$$

Interpretation: Je grösser s , desto stärker streuen die Werte um den Mittelwert.

4 Quartile und Interquartilsabstand

Quartile teilen sortierte Daten in vier Bereiche.

Kennwert	Bedeutung
Q_1	25% der Werte liegen darunter
Q_2	Median, 50% der Werte liegen darunter
Q_3	75% der Werte liegen darunter

Der Interquartilsabstand beschreibt die Streuung der mittleren 50% der Daten:

$$IQR = Q_3 - Q_1$$

Wichtig: Der IQR ist robuster gegen Ausreisser als die Spannweite oder Standardabweichung.

5 Beispiel

Gegeben sind die folgenden Daten:

$$4, 5, 7, 8, 8, 10, 12, 18$$

Lage

$$n = 8$$

$$\bar{x} = \frac{4 + 5 + 7 + 8 + 8 + 10 + 12 + 18}{8} = \frac{72}{8} = 9.00$$

Da es 8 Werte gibt, ist der Median der Durchschnitt des 4. und 5. Wertes:

$$\text{Median} = \frac{8 + 8}{2} = 8.00$$

Der Modus ist:

$$\text{Modus} = 8$$

Streuung

$$\text{Minimum} = 4 \quad \text{Maximum} = 18$$

$$\text{Spannweite} = 18 - 4 = 14$$

Für die Quartile werden die sortierten Daten in eine untere und obere Hälfte geteilt:

$$4, 5, 7, 8 \quad 8, 10, 12, 18$$

$$Q_1 = \frac{5 + 7}{2} = 6 \quad Q_3 = \frac{10 + 12}{2} = 11$$

$$IQR = Q_3 - Q_1 = 11 - 6 = 5$$

Varianz und Standardabweichung

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
4	-5	25
5	-4	16
7	-2	4
8	-1	1
8	-1	1
10	1	1
12	3	9
18	9	81
$\sum (x_i - \bar{x})^2$		138

$$s^2 = \frac{138}{8-1} = 19.71$$

$$s = \sqrt{19.71} = 4.44$$

6 Welche Kennwerte sind wichtig?

Situation	Lage	Streuung
keine starken Ausreisser	Mittelwert	Standardabweichung
starke Ausreisser	Median	IQR
kleine schnelle Beschreibung	Minimum, Maximum	Spannweite
kategoriale Daten	Modus	Häufigkeiten

Merksatz: Mittelwert und Standardabweichung passen gut zu ungefähr symmetrischen Daten. Median und IQR sind oft besser, wenn Ausreisser vorhanden sind.

7 Aufgaben

Aufgabe 1: Lagekennwerte

Berechnen Sie Mittelwert, Median und Modus für:

$$12, 15, 15, 18, 20, 24$$

Aufgabe 2: Quartile und IQR

Bestimmen Sie Q_1 , Median, Q_3 und IQR für:

$$3, 5, 7, 9, 10, 13, 15, 18, 22$$

Aufgabe 3: Streuung

Berechnen Sie Spannweite, Stichprobenvarianz und Standardabweichung für:

$$6, 8, 10, 12$$

Aufgabe 4: Ausreisser

Vergleichen Sie die beiden Datensätze:

$$A : 4, 5, 6, 7, 8$$

$$B : 4, 5, 6, 7, 28$$

Welcher Lagekennwert beschreibt Datensatz B besser: Mittelwert oder Median? Begründen Sie kurz.

Aufgabe 5: Multiple Choice

Kreuzen Sie alle richtigen Aussagen an.

- Der Median ist robust gegen Ausreisser.
- Die Spannweite verwendet nur Minimum und Maximum.
- Die Standardabweichung kann negativ sein.
- Der IQR beschreibt die Streuung der mittleren 50% der Daten.

8 Lösungen

Lösung 1

$$\bar{x} = \frac{12 + 15 + 15 + 18 + 20 + 24}{6} = \frac{104}{6} = 17.33$$

$$\text{Median} = \frac{15 + 18}{2} = 16.50$$

$$\text{Modus} = 15$$

Lösung 2

Die Daten sind bereits sortiert:

$$3, 5, 7, 9, 10, 13, 15, 18, 22$$

$$\text{Median} = 10$$

Untere Hälfte:

$$3, 5, 7, 9 \quad Q_1 = \frac{5 + 7}{2} = 6$$

Obere Hälfte:

$$13, 15, 18, 22 \quad Q_3 = \frac{15 + 18}{2} = 16.50$$

$$IQR = 16.50 - 6 = 10.50$$

Lösung 3

$$\bar{x} = \frac{6 + 8 + 10 + 12}{4} = 9$$

$$\text{Spannweite} = 12 - 6 = 6$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
6	-3	9
8	-1	1
10	1	1
12	3	9
$\sum(x_i - \bar{x})^2$		20

$$s^2 = \frac{20}{4 - 1} = 6.67$$

$$s = \sqrt{6.67} = 2.58$$

Lösung 4

Datensatz A:

$$\bar{x} = 6 \quad \text{Median} = 6$$

Datensatz B:

$$\bar{x} = \frac{4 + 5 + 6 + 7 + 28}{5} = 10 \quad \text{Median} = 6$$

Antwort: Für Datensatz B beschreibt der Median die typische Lage besser, weil der Ausreisser 28 den Mittelwert stark nach oben zieht.

Lösung 5

- ✓ Richtig – Der Median wird durch extreme Werte weniger stark beeinflusst.
- ✓ Richtig – Die Spannweite ist Maximum minus Minimum.
- ✗ Falsch – Die Standardabweichung ist nie negativ.
- ✓ Richtig – Der IQR ist $Q_3 - Q_1$ und beschreibt die mittleren 50%.