

Für die **Spannweite R** gilt:

$$R = x_{\max} - x_{\min}$$

x_{\max} größter Beobachtungswert
 x_{\min} kleinster Beobachtungswert

Die Spannweite gibt die Länge des Bereichs an, über den sich die Beobachtungswerte verteilen. Eine größere Spannweite bedeutet eine größere Streuung, aber da die Spannweite nur von zwei Werten (dem kleinsten und dem größten Wert, Ausreisser!) abhängt, ist die Aussagekraft relativ gering.

Der **Zentralwert x_{Med} (Median)** ist derjenige Wert, der in der Mitte steht, wenn alle Beobachtungswerte x_i der Größe nach geordnet sind.

Für eine **ungerade Anzahl n** von Beobachtungswerten gilt:

$$x_{\text{Med}} = x_{\frac{n+1}{2}}$$

Für eine **gerade Anzahl n** von Beobachtungswerten gilt:

$$x_{\text{Med}} = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

Die **Varianz s^2** ist das arithmetische Mittel der quadrierten Abweichungen der Beobachtungswerte x_i vom Mittelwert \bar{x} :

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

Für die **Standardabweichung s** gilt: $s = \sqrt{\text{Varianz}}$

Varianz und Standardabweichung können wie der Mittelwert \bar{x} nur bei **metrisch skalierten Merkmalen** bestimmt werden.

Berechnung der Varianz aus einer Häufigkeitstabelle:

$$s^2 = \frac{1}{n} \sum_{i=1}^r (x_i - \bar{x})^2 \cdot n_i = \sum_{i=1}^r (x_i - \bar{x})^2 \cdot h_i$$

Berechnung der Varianz aus einer klassierten Häufigkeitstabelle:

$$s^2 = \frac{1}{n} \sum_{i=1}^r (m_i - \bar{x})^2 \cdot n_i$$

\bar{x} Mittelwert
 x_i i-ter Beobachtungswert
 n Anzahl der Merkmalsträger
 n_i Häufigkeit für die Merkmalsausprägung x_i
bzw. für die i-te Klasse
 r Anzahl der Merkmalsausprägungen
bzw. der Klassen
 m_i Klassenmitte der i-ten Klasse

Der **Median** teilt einen **der Größe nach geordneten Datensatz** in der Mitte.

Die **Quartile** unterteilen diese beiden Hälften jeweils wieder in zwei gleich große Teile, so dass man vier gleich große Bereiche erhält.

Das **1. Quartil (= unteres Quartil q_u)** ist ein Wert, der **größer als 25 %** aller geordneten Beobachtungswerte ist.

Das **2. Quartil** heißt in der Regel **Median** oder **Zentralwert** und ist ein Wert, der **größer als 50 %** aller geordneten Beobachtungswerte ist.

Das **3. Quartil (= oberes Quartil q_o)** ist ein Wert, der **größer als 75 %** aller geordneten Beobachtungswerte ist.

Die Berechnungsformel für die Quartile hängt davon ab, ob das Ergebnis der Multiplikation von $\frac{1}{4}$ bzw. $\frac{1}{2}$ bzw. $\frac{3}{4}$ mit n ganzzahlig ist oder nicht. Ist das Ergebnis nicht ganzzahlig, so nimmt man den Wert des nächst höheren Rangplatzes, sonst den Mittelwert aus dem Wert dieses und des nächst höheren Rangplatzes. Die Quartile entsprechen also nicht zwangsläufig einem Wert aus dem Datensatz.

Der Quartilsabstand $q = q_o - q_u$ beschreibt die Länge des Bereichs der mittleren **50 %** der Beobachtungswerte. Er ist damit ein Maß für die Dichte der Daten in der zentralen Hälfte.