Streuungswerte:

1) Range (R) ⇨ ab metrischem Messniveau
2) Quartilabstand (QA) und mittlere Quartilabstand (MQA) ⇨ ab metrischem Messniveau
3) Durchschnittliche Abweichung (AD) ⇨ ab metrischem Messniveau
4) Varianz (s^2)⇨ ab metrischem Messniveau
5) Standardabweichung (s) ⇨ ab metrischem Messniveau
6) Variationskoeffizient (Vk) ⇨ ab ratioskaliertem Messniveau

1) Range (R):

- wird auch Spannweite bzw. Variationsbreite genannt
- ist definiert als die Differenz zwischen dem größten und kleinsten Messwert einer Verteilung:

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

Vorteil des Streuungsmaßes:
- R ist sehr einfach zu berechnen

Nachteil des Streuungsmaßes:
- R gibt keine Auskunft über die Streuung der übrigen Messwerte, die zwischen den Extremwerten liegen
- R wird von Extremwerten beeinflusst

Beispiel:

Verteilung A	Verteilung B
x_i	x_i
2	**2**
6	6
8	8
10	10
12	12
14	14
16	**99**
R = 16 - 2 = 14	**R = 99 - 2 = 97**

2) Der Quartilabstand (QA) und der mittlere Quartilabstand (MQA):

- Es gibt zwei weitere Streuungsmaße, die erheblich stabiler sind als der Range, weil sie nicht von Extremwerten der Verteilung abhängen.

- Das erste Maß ist der Quartilabstand (QA), definiert als:

  Quartilabstand (QA) = $Q_3 - Q_1$

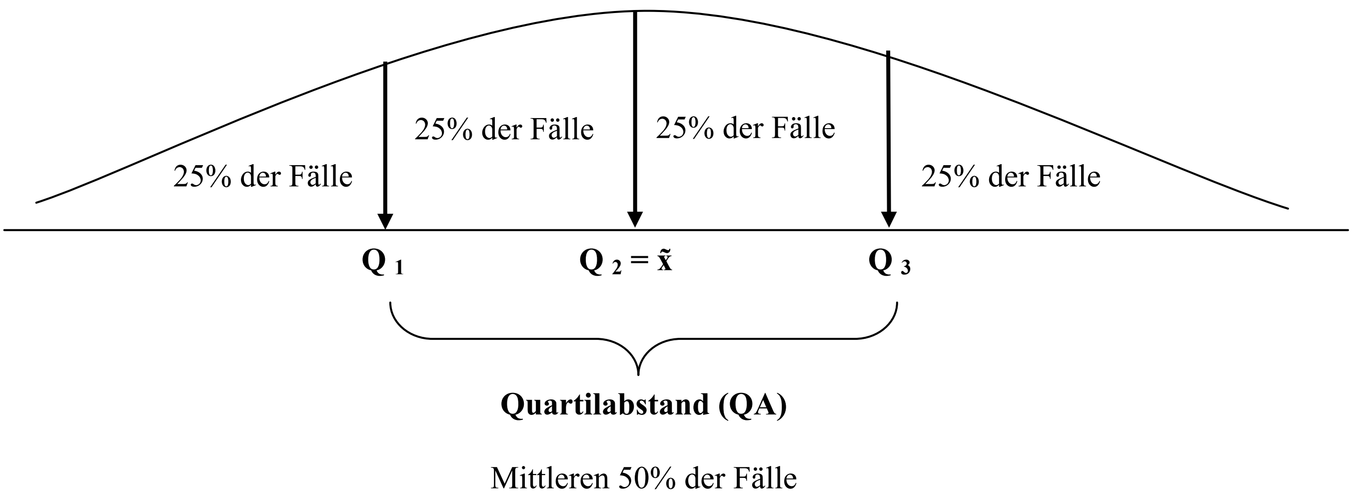
- Das zweite Maß ist der mittlere Quartilabstand (MQA), definiert als:

$$MQA = \frac{Q_3 - Q_1}{2}$$

- wobei Q_1 und Q_3 Quartile sind, nämlich das erste und dritte Quartil, die Schnittpunkte zwischen Vierteln der Verteilung bilden.
- Die Quartile trennen die unteren und oberen 25% der Fälle einer Verteilung von den mittleren 50% der Fälle.
- Einen weiteren Schnittpunkt bildet das zweite Quartil (Q_2), das die Verteilung halbiert und mit dem Median identisch ist.

- Der Quartilabstand ist demnach die Länge des Intervalls, das die mittleren 50% Fälle einer Beobachtungsreihe umfasst.

Illustration der Quartile und des Quartilabstandes:



Berechnung der Quartile:

Man ermittelt die Quartile in direkter Analogie zur Bestimmung des Medians.

Vorgehensweise:

1) Man ermittelt ¼ · N bzw. ¾ · N, d.h. die Anzahl der Fälle, die unterhalb Q_1 bzw. Q_3 liegen.

Q_1 = Messwert, unterhalb dessen genau ¼ und oberhalb dessen ¾ der Messwerte einer geordneten Reihe liegt.

$$Q_1 = \tfrac{1}{4} \cdot N$$

Q_3 = Messwert, unterhalb dessen genau ¾ und oberhalb dessen genau ¼ der Messwerte einer geordneten Reihe liegt.

$$Q_3 = \tfrac{3}{4} \cdot N$$

2) Man bestimmt anhand der kumulierten Häufigkeitsverteilung die (Klassen-) Intervalle, in die die Quartile Q_1 bzw. Q_3 fallen (Quartilintervalle). Das sind die Messwerte bzw. Intervalle mit einer kumulierten Häufigkeit **gleich** oder **(nächst) größer** als ¼ · N bzw. ¾ · N.
   - Ist die kumulierte Häufigkeit eines Intervalls genau gleich ¼ N, dann ist die exakte obere Grenze dieses Intervalls der Wert des Quartils Q_1; ist sie genau gleich ¾ N, dann ist die exakte obere Grenze dieses Intervalls der Wert des Quartils Q_3.

3) Man vergewissert sich der exakten unteren und oberen Grenzen der Quartilintervalle.

4) Man berechnet die Quartile nun nach der Formel:

$$Q_1 = U + \left[\frac{\tfrac{1}{4} \cdot N - Fu}{Fm}\right] \cdot h \qquad \text{und} \qquad Q_3 = U + \left[\frac{\tfrac{3}{4} \cdot N - Fu}{Fm}\right] \cdot h$$

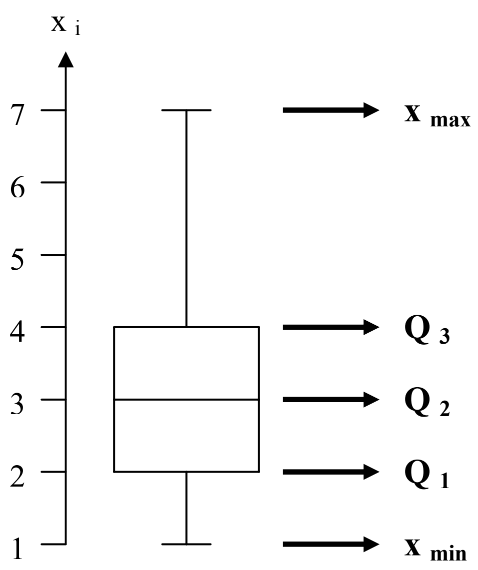
U = exakte untere Grenze des Quartilintervalls
N = Anzahl der Fälle
Fu = kumulierte Häufigkeit unterhalb des Quartilintervalls
Fm = Häufigkeit im Quartilintervall
h bzw. Kb = Breite des Quartilintervalls

- Die Quartile werden nicht durch Extremwerte beeinflusst.
- Eine mögliche **graphische Darstellung** der Quartile ist der **Boxplot**.

Konstruktion eines Boxplots:

Gegeben sei:

$Q_1 = 2$
$Q_2 = \tilde{x} = 3$
$Q_3 = 4$
$x_{min} = 1$
$x_{max} = 7$



5) Die durchschnittliche Abweichung (AD):

- Statistisches Streuungsmaß (wie die Varianz und Standardabweichung (siehe unten)), das die Verteilung der Messwerten um ihr arithmetische Mittel charakterisiert.

- Da die Summe der Abweichungen der Messwerte von ihrem arithmetischen Mittel immer gleich Null ist, müssen die negativen Vorzeichen ausgeschaltet werden.
- Die Durchschnittliche Abweichung, Standardabweichung und Varianz stellen verschiedene Versionen dar, die anfallenden negativen Vorzeichen bei der Errechnung der Abweichungen $\sum(x_i - \bar{x})$ zu umgehen.

- Die Differenzen der Messwerte zu ihrem arithmetischen Mittel werden hier mit Betrags-klammern absolut gesetzt.

$$AD = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N}$$

Beispiel:

Variable „Alter“				
x_i	f_i	$f_i x_i$	$\|x_i - \bar{x}\|$	$f_i \cdot \|x_i - \bar{x}\|$
1	1	1	|1-5| = 4	1 · 4 = 4
3	1	3	|3-5| = 2	1 · 2 = 2
6	1	6	|6-5| = 1	1 · 1 = 1
7	1	7	|7-5| = 2	1 · 2 = 2
8	1	8	|8-5| = 3	1 · 3 = 3
∑	N = 5	25	12	12

$$\bar{x} = \frac{25}{5} = 5$$

$$AD = \frac{12}{5} = 2{,}4$$

Interpretation:

- Ein AD-Wert von 2,4 besagt, dass die Messwerte im Durchschnitt 2,4 **Einheiten** von ihrem arithmetischen Mittel abweichen.

Auf die Variable „Alter“ bezogen bedeutet dies:
- Die Messwerte weichen durchschnittlich um 2,4 **Jahre** vom Altersdurchschnitt ($\bar{x}$ = 5 Jahre) ab.

6) Varianz (s^2):

- ist die Summe der quadrierten Abweichungen aller Messwerte von ihrem arithmetischen Mittel, geteilt durch ihre Anzahl:

$$s^2 = \frac{\sum f_i \cdot (x_i - \bar{x})^2}{N}$$

<u>Beispiel</u>:

Variable „Alter“					
x_i	f_i	$f_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i \cdot (x_i - \bar{x})^2$
1	1	1	1-5 = -4	16	1 · 16 = 16
3	1	3	3-5 = -2	4	1 · 4 = 4
6	1	6	6-5 = 1	1	1 · 1 = 1
7	1	7	7-5 = 2	4	1 · 4 = 4
8	1	8	8-5 = 3	9	1 · 9 = 9
$\sum$	N = 5	25	0	34	34

$$\bar{x} = \frac{25}{5} = 5$$

$$s^2 = \frac{34}{5} = 6{,}8$$

Interpretation:

- Ein s^2-Wert von 6,8 besagt, dass die Messwerte im Durchschnitt 6,8 **Quadrat-Einheiten** von ihrem arithmetischen Mittel abweichen.

Auf die Variable „Alter“ bezogen bedeutet dies:

- Die Messwerte weichen durchschnittlich um 6,8 **Quadrat-Jahre** vom Altersdurchschnitt ($\bar{x}$ = 5 Jahre) ab.

7) Standardabweichung (s):

- ist die Wurzel aus der Varianz:

$s = \sqrt{s^2}$ bzw. $s = \sqrt{\frac{\sum f_i \cdot (x_i - \overline{x})^2}{N}}$

<u>Beispiel</u>:

$s^2 = 6{,}8$. Demnach beträgt die Standardabweichung $s = \sqrt{6{,}8} = 2{,}61$

Interpretation:

- Ein s-Wert von 2,61 besagt, dass die Messwerte im Durchschnitt 2,61 **Einheiten** von ihrem arithmetischen Mittel abweichen.

Auf die Variable „Alter“ bezogen bedeutet dies:

- Die Messwerte weichen durchschnittlich um 2,61 **Jahre** vom Altersdurchschnitt ($\overline{x}$ = 5 Jahre) ab.

Allgemein:

- Standardabweichung und Varianz sind grundsätzlich als gleichwertige Streuungsmaße anzusehen, denn wenn die Varianz groß (klein) ist, ist auch die Standardabweichung groß (klein).
- Für *deskriptive Zwecke* ist allerdings die Standardabweichung vorzuziehen, weil sie ein Kennwert in der Einheit der zugrunde liegenden Messwerte ist (in dem hier vorliegenden Rechenbeispiel „Jahre“, nicht „Quadrat-Jahre“)

Variationskoeffizient (Vk, V) bzw. Ungleichheitsmaß:

Formel:

$$Vk = \frac{s}{\bar{x}} \quad , wobei\ \bar{x} > 0$$

- Der Varianzkoeffizient relativiert die Standardabweichung am Mittelwert. Der Variationskoeffizient drückt die Standardabweichung in Mittelwertseinheiten aus.
- Dieses Maß wird gelegentlich eingesetzt, wenn *Streuungen von Verteilungen mit unterschiedlichen Mittelwerten zu vergleichen* sind und Mittelwert und Streuung voneinander abhängen.

Beispiel: Haushaltjahreseinkommen in den Ländern A und B

Land A:	50.000	50.000	50.000
Land B:	1.000	1.000	148.000

$$\bar{x}_A = \bar{x}_B = 50.000,\ da\ \frac{150.000}{3} = 50.000$$

Land A				Land B			
x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i \cdot (x_i - \bar{x})^2$	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i \cdot (x_i - \bar{x})^2$
50.000	0	0	0	1.000	-49.000	2.401.000.000	2.401.000.000
50.000	0	0	0	1.000	-49.000	2.401.000.000	2.401.000.000
50.000	0	0	0	148.000	98.000	9.604.000.000	9.604.000.000
$\sum$	0	0	0	$\sum$	0	14.406.000.000	14.406.000.000

$$s_A = \sqrt{\frac{0}{3}} = \sqrt{0} = 0 \qquad \text{und} \qquad s_B = \sqrt{\frac{14.406.000.000}{3}} = \sqrt{4.802.000.000} = 69.296$$

- Die Standardabweichung s bezieht sich auf die Dimensionen der Messwerte, z.B. auf das Einkommen in Euro
- Der Variationskoeffizient ist eine dimensionslose Größe und unempfindlich gegenüber linearen Transformationen (z.B. Wechselkursumrechnung)

$$Vk_A = \frac{0}{50.000} = 0 \qquad \text{und} \qquad Vk_B = \frac{69.296}{50.000} = 1{,}39$$

Interpretation:

- In beiden Ländern streut, gemessen am Durchschnitt, das Haushaltseinkommen ungleich. Die relative Streuung ist für das Land B größer als für das Land A.