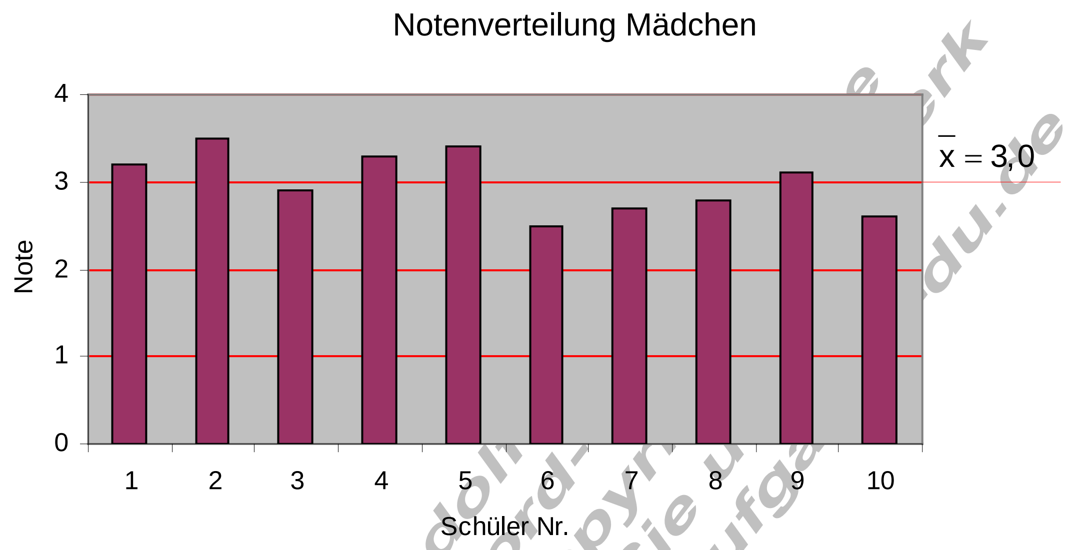
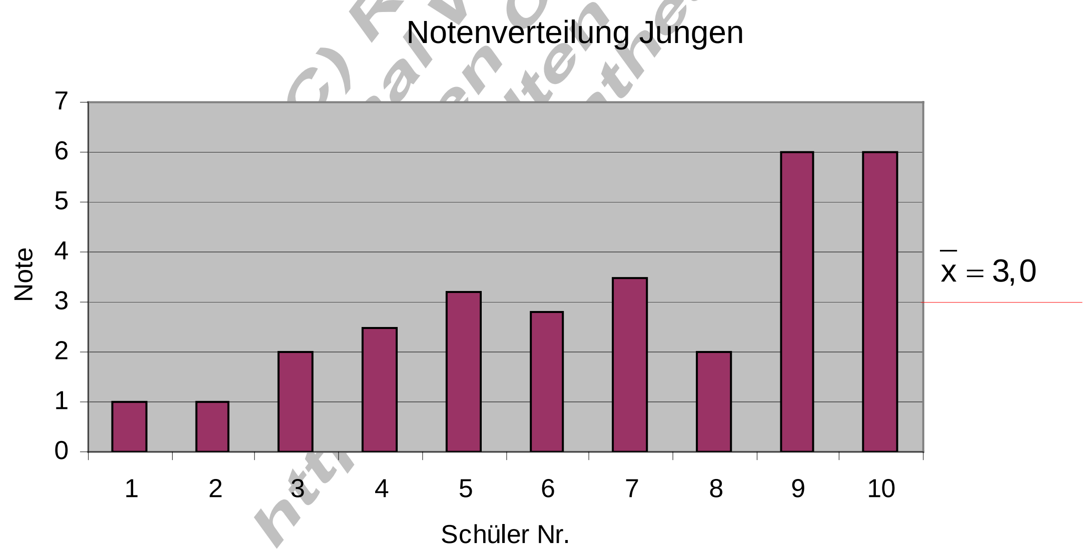
Spannweite, Median Quartilsabstand, Varianz und Standardabweichung.

<u>Streuung um den Mittelwert.</u>

In den folgenden Säulendiagrammen ist die Notenverteilung zweier Schülergruppen (Mädchen, Jungen) dargestellt, deren Mittelwert gleich ist.

Schüler Nr.	1	2	3	4	5	6	7	8	9	10	
Note Mädchen	3,2	3,5	2,9	3,3	3,4	2,5	2,7	2,8	3,1	2,6	$\overline{x} = 3{,}0$
Note Jungen	1,0	1,0	2,0	2,5	3,2	2,8	3,5	2,0	6,0	6,0	$\overline{x} = 3{,}0$





Bei den Mädchen liegen die Noten alle sehr nahe am Mittelwert
Sie streuen wenig um den Mittelwert.
Bei den Jungen sind die Abweichungen vom Mittelwert sehr groß.
Sie streuen stark um den Mittelwert.
Die Statistik bietet Möglichkeiten, die Streuung näher zu untersuchen.

Die Spannweite.

Berechnet man den Unterschied zwischen dem größten und kleinsten Beobachtungswert, so erhält man die Spannweite.
Sie ist ein Maß für die Breite des Streubereichs einer Häufigkeitsverteilung.

Die Spannweite
Spannweite = größter Beobachtungswert - kleinster Beobachtungswert
$R = x_{max} - x_{min}$

Beispiel:

Schüler Nr.	1	2	3	4	5	6	7	8	9	10	
Note Mädchen	3,2	3,5	2,9	3,3	3,4	2,5	2,7	2,8	3,1	2,6	$\overline{x} = 3{,}0$
Note Jungen	1,0	1,0	2,0	2,5	3,2	2,8	3,5	2,0	6,0	6,0	$\overline{x} = 3{,}0$

Spannweite Mädchen: $R_M = 3{,}5 - 2{,}5 = \underline{\underline{1}}$

Spannweite Jungen: $R_J = 6{,}0 - 1{,}0 = \underline{\underline{5}}$

Der Quartilsabstand.

Zur Erinnerung: Der Median teilt einen nach Größe sortierten Datensatz in der Mitte.
Das bedeutet, links und rechts vom Median liegen gleich viele Beobachtungswerte.

$$\frac{\text{höchstens 50\% aller B - Werte} \leq \text{Median} \leq \text{höchstens 50\% aller B - Werte}}{\text{links vom Median} \leq \text{Median} \leq \text{rechts vom Median}}$$

Unterteilt man die linke und die rechte Hälfte nach gleicher Vorschrift, wie man den Median bestimmt, so erhält man vier gleich große Bereiche, die durch drei Quartile aufgeteilt werden.

Beispiel:
Die Liste enthält von 13 Schülern die Körpergröße.
Die Merkmalsausprägungen (Beobachtungswerte) wurden nach der Größe geordnet.

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
KG	1,60	1,67	1,67	1,68	1,68	1,70	1,70	1,72	1,73	1,75	1,76	1,78	1,84

$x_i =$ Beobachtungswert x_i ; KG = Körpergröße in m

Median / 2. Quartil: $Q_2 = x_7 = \underline{\underline{1{,}70}}$

1. Quartil: $Q_1 = \frac{1}{2}(x_3 + x_4) = \frac{1}{2}(1{,}67 + 1{,}68) = \underline{\underline{1{,}675}}$

3. Quartil: $Q_3 = \frac{1}{2}(x_{10} + x_{11}) = \frac{1}{2}(1{,}75 + 1{,}76) = \underline{\underline{1{,}755}}$

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
KG	1,60	1,67	**1,67**	**1,68**	1,68	1,70	**1,70**	1,72	1,73	**1,75**	**1,76**	1,78	1,84

25% 25% 25% 25%

1. Quartil $Q_1 = 1{,}675$ 2. Quartil $Q_2 = 1{,}70$ 3. Quartil $Q_3 = 1{,}755$

50%

Quartilsabstand

Definition:	25% aller geordneten Beobachtungswerte sind kleiner als das 1. Quartil 50% aller geordneten Beobachtungswerte sind kleiner als das 2. Quartil. 75% aller geordneten Beobachtungswerte sind kleiner als das 3. Quartil.

Wie leicht zu erkennen ist, liegen zwischen dem 1. und 3. Quartiel 50% aller Beobachtungswerte. Dieser Bereich wird auch Quartilsabstand genannt.

Quartilsabstand

Der mittlere 50% - Bereich aller Beobachtungswerte heißt Quartilsabstand.

Berechnung: $Q_A = Q_3 - Q_1$

Weitere Auswertung des Beispiels:

Quartilsabstand: $Q_A = Q_3 - Q_1 = 1{,}755 - 1{,}675 = 0{,}08$

50% der Daten liegen in einem Bereich der Bandbreite von 0,08 m bzw. 8 cm.

Etwa 50% der Körpergrößen liegen zwischen 1,675 m und 1,755 m.

Vergleich zwischen Quartilsabstand und Spannweite	
Quartilsabstand	Spannweite
Von Ausreißern unabhängig Gibt die Breite des mittleren Bereichs an, in dem ca. 50% aller Werte liegen	Vom kleinsten und größten Wert abhängig Gibt die Gesamtbreite an in dem alle Werte liegen

Beispiel:
Ein Landwirt misst im Monat April jeweils mittags um 12 Uhr die Außentemperatur und trägt sie in eine Tabelle ein.
Berechnen Sie den Mittelwert, Spannweite und Median.
Berechnen Sie das 1. und 3. Quartil und den Quartilsabstand.

Tag	1	2	3	4	5	6	7	8	9	10
Temperatur	7	10	12	16	16	17	18	20	22	29
Tag	11	12	13	14	15	16	17	18	19	20
Temperatur	23	19	20	21	18	17	15	29	22	23
Tag	21	22	23	24	25	26	27	28	29	30
Temperatur	8	25	24	23	23	25	26	27	19	16

Mittelwert: $\bar{x} = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{1}{30}(7+10+...+19+16) = \frac{590}{30} = \underline{\underline{19,\bar{6}}}$

```
0  7  8
1  0  2  5  6  6  6  7  7  8  8  9  9
2  0  0  1  2  2  3  3  3  3  4  5  5  6  7  9  9
```

Spannweite: $R = x_{max} - x_{min} = 29 - 7 = \underline{\underline{22}}$

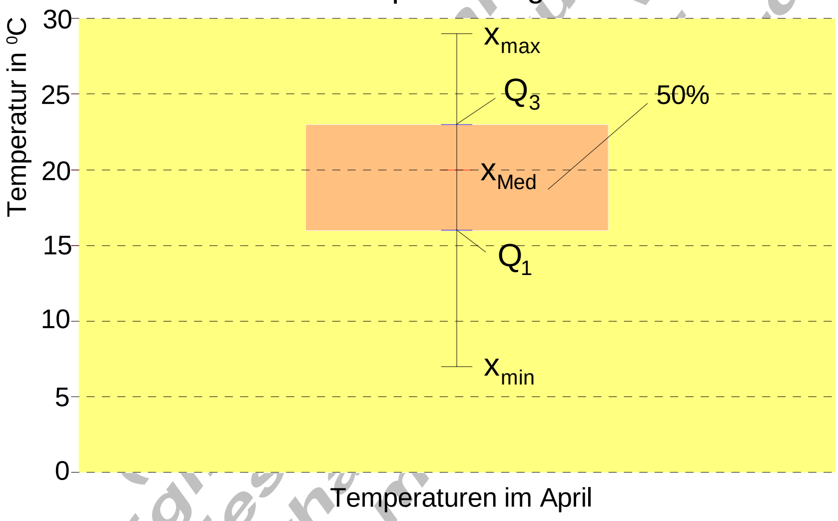
Median: $x_{Med} = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(20+20) = \underline{\underline{20}}$ (2. Quartil)

1. Quartil: $Q_1 = x_8 = \underline{\underline{16}}$

3. Quartil: $Q_3 = x_{23} = \underline{\underline{23}}$

Quartilsabstand: $Q_A = Q_3 - Q_1 = 23 - 16 = \underline{\underline{7}}$

Die Ergebnisse lassen sich in einem Boxplot – Diagramm darstellen:

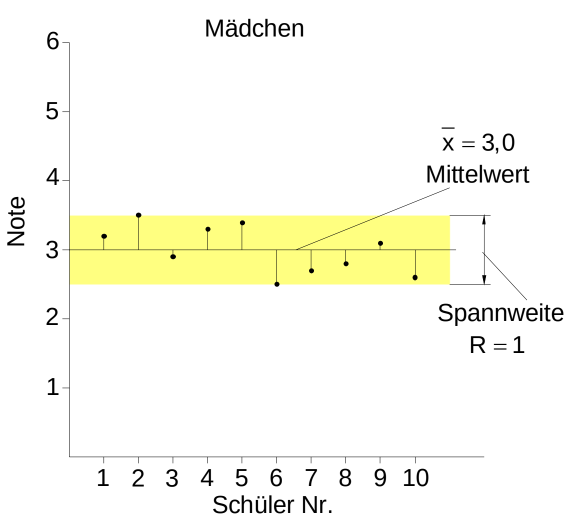


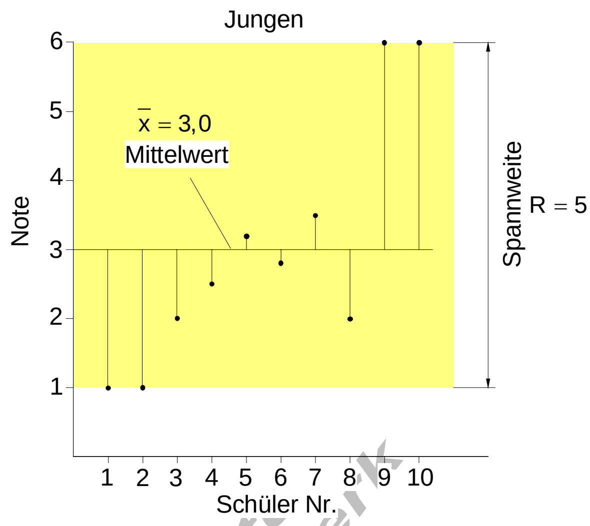
Varianz und Standardabweichung.

Wir betrachten noch mal die Notenverteilung von Mädchen und Jungen aus dem vorigen Beispiel.

Schüler Nr.	1	2	3	4	5	6	7	8	9	10	
Note Mädchen	3,2	3,5	2,9	3,3	3,4	2,5	2,7	2,8	3,1	2,6	$\bar{x} = 3,0$
Note Jungen	1,0	1,0	2,0	2,5	3,2	2,8	3,5	2,0	6,0	6,0	$\bar{x} = 3,0$

Der Mittelwert ist in beiden Fällen gleich, die Streuung um diesen ist unterschiedlich.





Abweichung

Die beiden Diagramme veranschaulichen die Abweichungen vom Mittelwert.

Abwichung: $x_i - \bar{x}$

In der beschreibenden Statistik berechnet man das arithmetische Mittel der Abweichungsquadrate und nennt dieses die Varianz.

Varianz einer Datenreihe

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{\left(x_1 - \bar{x}\right)^2 + \left(x_2 - \bar{x}\right)^2 + \ldots + \left(x_n - \bar{x}\right)^2}{n}$$

n: Anzahl der Beobachtungswerte, x_i : i - ter Beobachtungswert, $\bar{x}$: Mittelwert

Für unser Beispiel gilt:

Mädchen				Jungen			
i	x_i	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2$	i	x_i	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2$
1	3,2	0,2	0,04	1	1,0	–2,0	4,0
2	3,5	0,5	0,25	2	1,0	–2,0	4,0
3	2,9	–0,1	0,01	3	2,0	–1,0	1,0
4	3,3	0,3	0,09	4	2,5	–0,5	0,25
5	3,4	0,4	0,16	5	3,2	0,2	0,04
6	2,5	–0,5	0,25	6	2,8	–0,2	0,04
7	2,7	–0,3	0,09	7	3,5	0,5	0,25
8	2,8	–0,2	0,04	8	2,0	–1,0	1,0
9	3,1	0,1	0,01	9	6,0	3,0	9,0
10	2,6	–0,4	0,16	10	6,0	3,0	9,0
$\sum$	30	0	1,10	$\sum$	30	0	28,58

Varianz Mädchen: $s_M^2 = \frac{1}{10} \cdot 1,1 = \underline{\underline{0,11}}$ Varianz Jungen: $s_J^2 = \frac{1}{10} \cdot 28,58 = \underline{\underline{2,858}}$

Viele Daten sind mit Einheiten behaftet, z.B. Meter (m) oder kg.
Die Einheit für die Varianz wäre in diesen Fällen m^2 bzw. $(kg)^2$.
Um wieder auf die ursprüngliche Einheit zu kommen, zieht man die Wurzel aus der Varianz. Dieser Wert wird Standardabweichung genannt.

Standardabweichung:	$s = \sqrt{s^2} = \sqrt{\text{Varianz}}$

Zur praktischen Berechnung fertigt man wie oben gezeigt eine entsprechende Tabelle an. Sie dient auch zur Kontrolle der Daten.
Die Summe der Abweichungen muss Null ergeben.

Bemerkung zur Varianz:
Handelt es sich bei den zu untersuchenden Daten um die Population (Grundgesamtheit), dann wird mit $1/n$ gewichtet:

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

Wird hingegen eine Stichprobe (Teil einer Population) untersucht, so wird mit $1/(n-1)$ gewichtet:

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

Berechnung der Standardabweichung aus einer Häufigkeitstabelle.

Hier geht man ähnlich vor wie bei der Mittelwertbildung.
Zur Erinnerung:

Fall I: Absolute Häufigkeit n_i

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{j} x_i \cdot n_i = \frac{1}{n}\left(x_1 \cdot n_1 + x_2 \cdot n_2 + ... + x_j \cdot n_j\right) \qquad n = \sum_{i=1}^{j} n_i = n_1 + n_2 + ... + n_j$$

Fall II: Relative Häufigkeit $h_i = \frac{n_i}{n}$

$$\overline{x} = \sum_{i=1}^{j} x_i \cdot h_i = \left(x_1 \cdot h_1 + x_2 \cdot h_2 + ... + x_j \cdot h_j\right)$$

n_i : absolute Häufigkeit der Merkmalsausprägung x_i
h_i : relative Häufigkeit der Merkmalsausprägung x_i
n : Summe der absoluten Häufigkeiten
j : Anzahl der Merkmalsausprägungen x_i

Berechnung der Varianz aus einer Häufigkeitstabelle

Fall I: Absolute Häufigkeit n_i

$$n = \sum_{i=1}^{j} n_i = n_1 + n_2 + \ldots + n_j$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{j}\left(x_i - \overline{x}\right)^2 \cdot n_i = \frac{\left(x_1 - \overline{x}\right)^2 \cdot n_1 + \left(x_2 - \overline{x}\right)^2 \cdot n_2 + \ldots + \left(x_j - \overline{x}\right)^2 \cdot n_j}{n}$$

Fall II: Relative Häufigkeit $h_i = \dfrac{n_i}{n}$

$$s^2 = \sum_{i=1}^{j}\left(x_i - \overline{x}\right)^2 \cdot h_i = \left(x_1 - \overline{x}\right)^2 \cdot h_1 + \left(x_2 - \overline{x}\right)^2 \cdot h_2 + \ldots + \left(x_j - \overline{x}\right)^2 \cdot h_j$$

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

n_i : absolute Häufigkeit der Merkmalsausprägung x_i
n : Summe der absoluten Häufigkeiten
h_i : relative Häufigkeit der Merkmalsausprägung x_i
j : Anzahl der Merkmalsausprägungen x_i

Beispiel:

Note (x_i)	1	2	3	4	5	6
Anz. d. Schüler (n_i)	5	8	14	16	5	2

Note (x_i)	1	2	3	4	5	6
Anz. d. Schüler (n_i)	5	8	14	16	5	2

Schüler insgesamt: $n = \sum_{i=1}^{6} n_i = 50$

i	x_i	n_i	$x_i \cdot n_i$	$\overline{x}$	$x_i - \overline{x}$	$\left(x_i - \overline{x}\right)^2 \cdot n_i$
1	1	5	5	3,28	−2,28	25,992
2	2	8	16	3,28	−1,28	13,1072
3	3	14	42	3,28	−0,28	1,0976
4	4	16	64	3,28	0,72	8,2944
5	5	5	25	3,28	1,72	14,792
6	6	2	12	3,28	2,72	14,7968
$\sum$		50	164	$\overline{x} = \frac{164}{50} = 3,28$		78,08

Varianz: $s^2 = \frac{1}{50}\sum_{i=1}^{6}\left(x_i - \overline{x}\right)^2 \cdot n_i = \frac{78,08}{50} = \underline{\underline{1,5616}}$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{1,5616} = \underline{\underline{1,2496}}$

Beispiel:

Note (x_i)	1	2	3	4	5	6
Anz. d. Schüler (n_i)	5	8	14	16	5	2
rel. Häufigkeit $h_i = \frac{n_i}{n}$	0,1	0,16	0,28	0,32	0,1	0,04

Schüler insgesamt: $n = \sum_{i=1}^{6} n_i = 50$

i	x_i	h_i	$x_i \cdot h_i$	$\overline{x}$	$x_i - \overline{x}$	$(x_i - \overline{x})^2 \cdot h_i$
1	1	0,1	0,1	3,28	−2,28	0,51984
2	2	0,16	0,32	3,28	−1,28	0,262144
3	3	0,28	0,84	3,28	−0,28	0,021952
4	4	0,32	1,28	3,28	0,72	0,165888
5	5	0,1	0,50	3,28	1,72	0,29584
6	6	0,04	0,24	3,28	2,72	0,295936
$\sum$		1	$\overline{x} = 3,28$			$s^2 = 1,5616$

Varianz: $s^2 = \sum_{i=1}^{6} (x_i - \overline{x})^2 \cdot h_i = \underline{\underline{1,5616}}$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{1,5616} = \underline{\underline{1,2496}}$

Das Beispiel zeigt, dass es sich mit den relativen Häufigkeiten leichter rechnen lässt.

Berechnung der Standardabweichung aus einer klassierten Häufigkeitstabelle.

Zur Erinnerung:

Fall I: Absolute Häufigkeit n_i

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} m_i \cdot n_i = \frac{1}{n}(m_1 \cdot n_1 + m_2 \cdot n_2 + ... + m_k \cdot n_k) \qquad n = \sum_{i=1}^{k} n_i = n_1 + n_2 + ... + n_k$$

Fall II: Relative Häufigkeit $h_i = \frac{n_i}{n}$

$$\overline{x} = \sum_{i=1}^{k} m_i \cdot h_i = m_1 \cdot h_1 + m_2 \cdot h_2 + ... + m_k \cdot h_k$$

n_i : absolute Häufigkeit der i - ten Klasse   n : Summe der absoluten Häufigkeiten
h_i : relative Häufigkeit der i - ten Klasse   k : Anzahl der Klassen
m_i : Klassenmitte der i -ten Klasse

Berechnung der Varianz aus einer klassierten Häufigkeitstabelle

Fall I: Absolute Häufigkeit n_i

$$n = \sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_k$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}\left(m_i - \overline{x}\right)^2 \cdot n_i = \frac{\left(m_1 - \overline{x}\right)^2 \cdot n_1 + \left(m_2 - \overline{x}\right)^2 \cdot n_2 + \ldots + \left(m_k - \overline{x}\right)^2 \cdot n_k}{n}$$

Fall II: Relative Häufigkeit $h_i = \frac{n_i}{n}$

$$s^2 = \sum_{i=1}^{k}\left(m_i - \overline{x}\right)^2 \cdot h_i = \left(m_1 - \overline{x}\right)^2 \cdot h_1 + \left(m_2 - \overline{x}\right)^2 \cdot h_2 + \ldots + \left(m_k - \overline{x}\right)^2 \cdot h_k$$

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

n_i : absolute Häufigkeit der i - ten Klasse
h_i : relative Häufigkeit der i - ten Klasse
m_i : Klassenmitte der i -ten Klasse
n : Summe der absoluten Häufigkeiten
k : Anzahl der Klassen

Beispiel :
Bestimmen Sie aus der klassierten Häufigkeitstabelle für die Körpergröße die Standardabweichung.

Klasse x_i	$150 \le x < 160$	$160 \le x < 170$	$170 \le x < 180$	$180 \le x < 190$
abs. Häufigkeit n_i	9	12	7	2
Klassenmitte m_i	155	165	175	185
rel. Häufigkeit $h_i = \frac{n_i}{n}$	$0,3$	$0,4$	$0,2\overline{3}$	$0,0\overline{6}$

$$\text{Klassenmitte} = \frac{\text{Klassenanfang} + \text{Klassenende}}{2} \quad \text{z.B.} \quad \frac{160+170}{2} = 165$$

Berechnung über die absolute Häufigkeit

i	m_i	n_i	$m_i \cdot n_i$	$\overline{x}$	$m_i - \overline{x}$	$\left(m_i - \overline{x}\right)^2 \cdot n_i$
1	155	9	1395	$165,\overline{6}$	$-10,\overline{6}$	$1023,\overline{9}$
2	165	12	1980	$165,\overline{6}$	$-0,\overline{6}$	$5,\overline{3}$
3	175	7	1225	$165,\overline{6}$	$9,\overline{3}$	$609,\overline{7}$
4	185	2	370	$165,\overline{6}$	$19,\overline{3}$	$747,\overline{5}$
$\sum$		$n = 30$	4970	$\frac{4970}{30} = 165,\overline{6}$		$2386,\overline{6}$

Varianz: $s^2 = \frac{2386,\overline{6}}{30} = \underline{\underline{79,\overline{5}}}$

Standardabweichung: $s = \sqrt{79,\overline{5}} \approx \underline{\underline{8,9194}}$

Berechnung über die relative Häufigkeit

i	m_i	h_i	$m_i \cdot h_i$	$\overline{x}$	$m_i - \overline{x}$	$(m_i - \overline{x})^2 \cdot h_i$
1	155	$0,3$	$46,5$	$165,\overline{6}$	$-10,\overline{6}$	$34,1\overline{3}$
2	165	$0,4$	$66,0$	$165,\overline{6}$	$-0,\overline{6}$	$0,1\overline{7}$
3	175	$0,2\overline{3}$	$40,8\overline{3}$	$165,\overline{6}$	$9,\overline{3}$	$20,3259$
4	185	$0,0\overline{6}$	$12,\overline{3}$	$165,\overline{6}$	$19,\overline{3}$	$24,91$
$\sum$		1	$165,\overline{6}$			$79,\overline{5}$

Varianz: $s^2 = \underline{\underline{79,\overline{5}}}$ Standardabweichung: $s = \sqrt{79,\overline{5}} \approx \underline{\underline{8,9194}}$

Auch hier lässt sich das Problem einfacher über die relative Häufigkeit lösen.

Die Standardabweichung ist ein Maß dafür, wie hoch die Aussagekraft des Mittelwertes ist.
Eine kleine Standardabweichung bedeutet, alle Beobachtungswerte liegen nahe am Mittelwert.
Eine große Standardabweichung bedeutet, die Beobachtungswerte sind weit um den Mittelwert gestreut.