Empirische Varianz am Beispiel des Notenspiegels von Klassenarbeiten

1	2	3	4	5	6
3	3	5	2	4	4

Die 21 Schülerinnen und Schüler haben eine Klassenarbeit abgeliefert,
deren Ausfall augenscheinlich nicht nur schlecht sondern auch breit gestreut ist.

Der Durchschnitt wird, wie allseits bekannt, wie folgt berechnet:

$$\overline{x} = (3 \cdot 1 + 3 \cdot 2 + 5 \cdot 3 + 2 \cdot 4 + 4 \cdot 5 + 4 \cdot 6) : 21 \approx 3{,}62$$

Aber warum macht man das so?

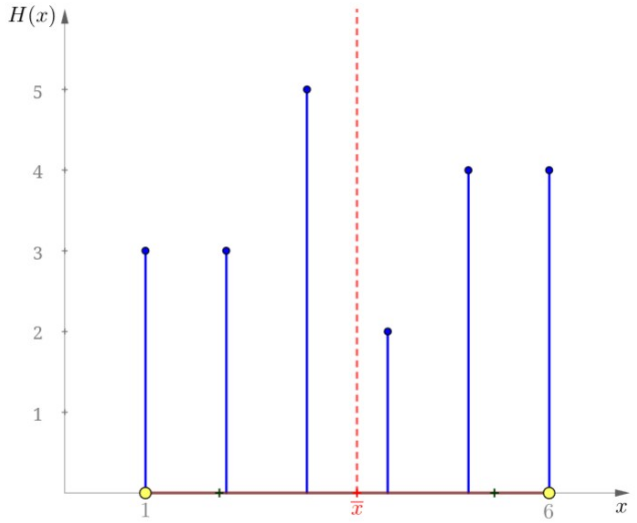
Die 21 Einzel-Ereignisse sortiert als x_i sehen wie folgt aus:

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}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}
1	1	1	2	2	2	3	3	3	3	3	4	4	5	5	5	5	6	6	6	6

Nun ist klar, dass der Mittelwert der 21 Ereignisse gemäß der obigen Rechnung zu bestimmen ist.

© *Thorsten Glaser, Gymnasium Trittau*

Als Stabdiagramm dargestellt wird die Breite der Streuung deutlich.
Doch wie ermittelt man einen Wert für das Maß der Streuung?



Ein erster Gedanke liegt nahe: Die Betrachtung der Differenzen zum Mittelwert.

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}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}
1	1	1	2	2	2	3	3	3	3	3	4	4	5	5	5	5	6	6	6	6
-2,62	-2,62	-2,62	-1,62	-1,62	-1,62	-0,62	-0,62	-0,62	-0,62	-0,62	0,38	0,38	1,38	1,38	1,38	1,38	2,38	2,38	2,38	2,38

Da die Differenzen sowohl positiv als auch negativ sind, kann ein Mittelwert daraus so noch nicht direkt berechnet werden.

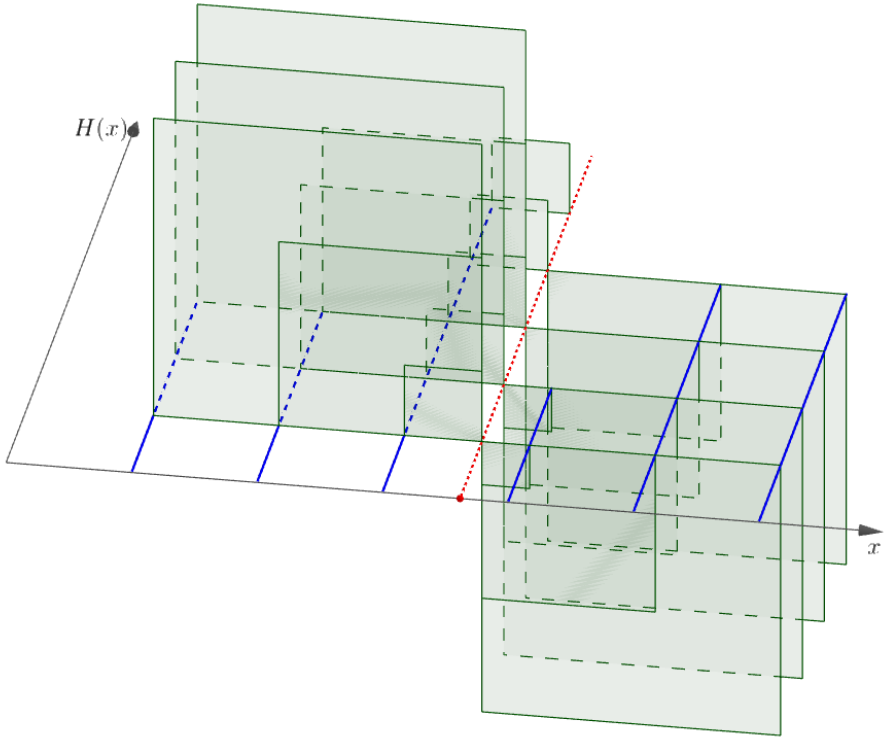
Statt die *Beträge* der Abweichungen zu verwenden, verwendete Carl Friedrich Gauß Abweichungs-*Quadrate*. Das noch heute allgemein so üblich.

© *Thorsten Glaser, Gymnasium Trittau*

In der vorletzten Zeile noch einmal die Differenzen zum Mittelwert,
in der untersten Zeile deren quadrierte Werte:

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}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}
1	1	1	2	2	2	3	3	3	3	3	4	4	5	5	5	5	6	6	6	6
-2,62	-2,62	-2,62	-1,62	-1,62	-1,62	-0,62	-0,62	-0,62	-0,62	-0,62	0,38	0,38	1,38	1,38	1,38	1,38	2,38	2,38	2,38	2,38
6,86	6,86	6,86	2,62	2,62	2,62	0,38	0,38	0,38	0,38	0,38	0,15	0,15	1,91	1,91	1,91	1,91	5,67	5,67	5,67	5,67

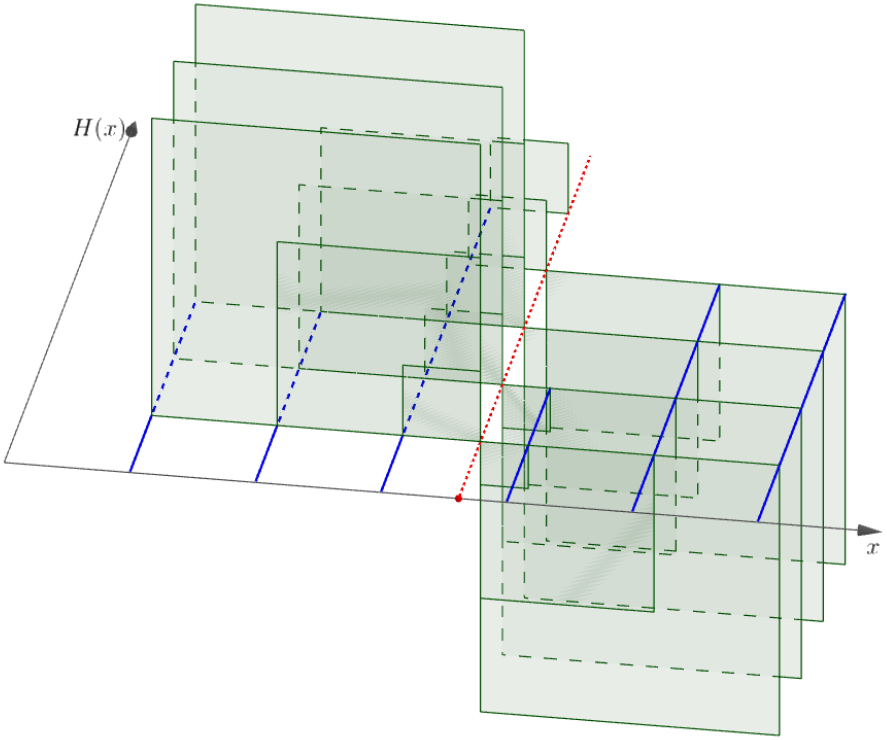
Die Abweichungsquadrate sind hier visualisiert im $\mathbb{R}^3$



© *Thorsten Glaser, Gymnasium Trittau*

Der Durchschnitt der Abweichungs-Quadrate ist definiert als die Varianz:

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



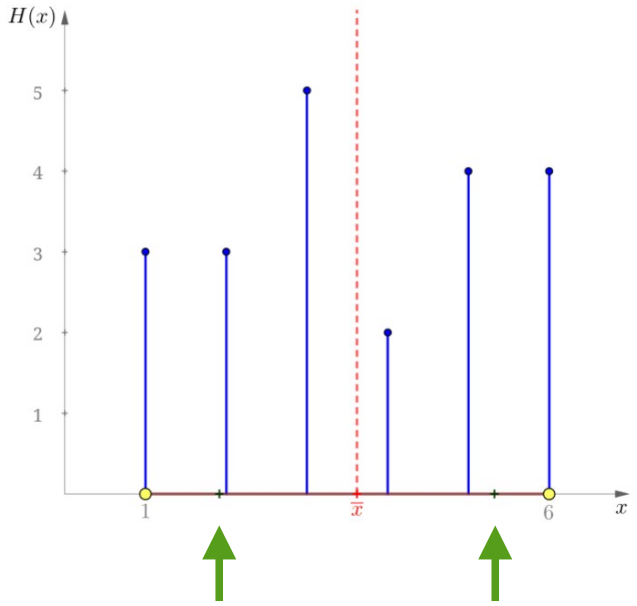
In der wissenschaftlichen Statistik teilt man nicht durch die Anzahl n, sondern durch die Anzahl der Freiheitsgrade n -1.

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

© *Thorsten Glaser, Gymnasium Trittau*

Um wieder auf ein Längenmaß zu kommen, zieht man die Wurzel aus der Varianz.

Dieses Streumaß nennt man die Standardabweichung: $s = \sqrt{s^2}$



Im Stabdiagramm sind die Stellen $\overline{x} - s$ und $\overline{x} + s$ markiert.

© *Thorsten Glaser, Gymnasium Trittau*