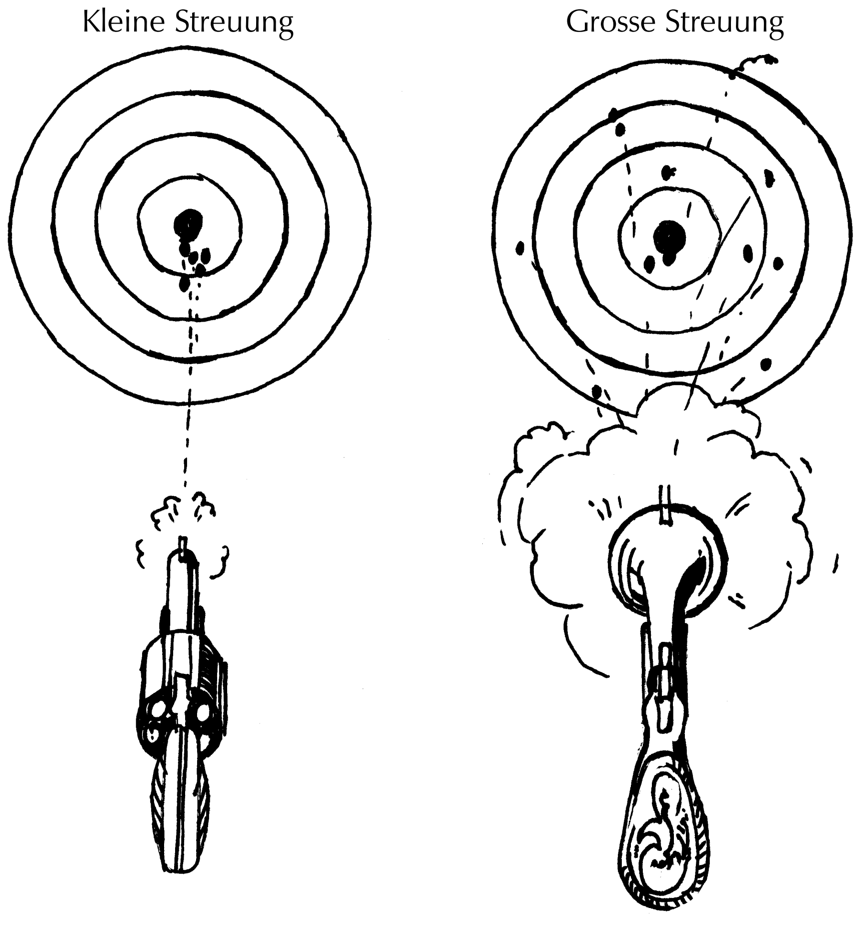
4. Streuungsmasse

4.1 Überblick

Streuungsmasse sagen etwas darüber aus, wie stark die Werte voneinander abweichen oder „streuen". Liegen die Werte nahe beisammen, so ist die Streuung klein; weichen die Werte stärker voneinander ab, so ist die Streuung gross. Zunächst einige Beispiele:



Prüfung I: Schüler A schreibt in einer Prüfung zweimal eine 5, Schüler B einmal eine 4 und einmal eine 6. Beide Schüler haben im Durchschnitt eine 5. Aber bei Schüler A ist die Streuung 0, bei Schüler B sicher grösser als Null. Wie gross die Streuung genau ist, hängt davon ab, mit welchem Streuungsmass sie gemessen wird.
Experimente: Messungen werden oft mehrfach unter denselben Versuchsbedingungen durchgeführt. Das Streuungsmass sollte deshalb möglichst 0 sein. Dies ist es in aller Regel aber nicht, weil die einzelnen Messresultate leicht von einander abweichen. Deshalb gibt das Streuungsmass Auskunft z.B. über die Mess(un)genauigkeit oder den Verfahrensfehler. Eine allzu grosse Streuung wirft deshalb die Frage auf, wie geeignet der durchgeführte Versuch überhaupt ist.
Prüfung II: In welcher der drei Klassen A, B und C variieren die Noten am wenigsten und ist daher die Streuung am kleinsten?

Prüfung			
Note	Häufigkeit		
	Klasse A	Klasse B	Klasse C
6	II		II
5½	III		IIIII I
5	IIII	IIIII	III
4½	IIIII	III	III
4	III	IIII	II
3½	II	III	II
3	III	I	
2½	II		
2			I
1½			I
1	I		

Sicher bei Klasse B, dort liegen die Noten in einem viel kleineren Bereich als bei den beiden anderen Klassen. Weil die Noten relativ wenig voneinander abweichen, ist die Streuung vergleichsweise klein – wie auch immer sie gemessen wird.
Bei den Klassen A und C dürften die Ausreisser nach unten für eine grosse Streuung sorgen. Aber wo ist sie grösser? Die Antwort hängt davon ab, mit welchem Streuungsmass gemessen wird.
Wir untersuchen nun einige Streuungsmasse und ihre Eigenschaften.

4.2 Einige Streuungsmasse

4.2.1 Die Spannweite (Variationsbreite)

Eine naheliegende Idee ist, als Streuungsmass die Differenz zwischen dem grössten und dem kleinsten Wert anzugeben: $x_{max} - x_{min}$. Dieses Streuungsmass nennt man Spannweite oder Variationsbreite.

Spannweite (Variationsbreite) $= x_{max} - x_{min}$ **[4.1]**

Sie ist extrem empfindlich auf Ausreisser und wird deshalb nicht häufig verwendet.
Im Standardbeispiel „Prüfung“ sind die Spannweiten für Klasse A 6–1=5, für Klasse B dagegen 5–3=2, für Klasse C schliesslich 6–1.5=4.5. Man erkennt bei den Klassen A und C deutlich den Einfluss der Ausreisser: Ohne diese wären die Spannweiten nicht 5 bzw. 4.5, sondern lediglich 3.5 bzw. 2.5.

4.2.2 Die Summe der Abweichungen vom Mittelwert

Eine weitere naheliegende Idee ist es, die Abweichungen vom Mittelwert zusammenzuzählen. Je grösser die Summe dieser Abweichungen ist, desto grösser die Streuung. Diese Idee hat zwei Nachteile:

- Erstens spielt die Anzahl der untersuchten Werte eine Rolle. Im Beispiel *Prüfung* bedeutet dies, dass bei kleinen Klassen eher kleine Streuungen zu erwarten sind, bei grossen Klassen eher grosse. Dies deshalb, weil bei einer grossen Klasse mehr Abweichungen addiert werden als bei einer kleineren – und zwar auch dann, wenn die Noten bei der grossen Klasse näher beisammen liegen als bei einer kleineren.

- Und zweitens gibt es einen mathematischen Grund, sobald man als Lagemass den Mittelwert verwendet, was der Normalfall ist. Die Summe der Abweichungen ist 0, und zwar gemäss Formel [3.6] immer. Deshalb wäre die Streuung bei Verwendung dieses Masses immer 0, und das ist sicher nicht besonders sinnvoll...

4.2.3 Die mittlere absolute Abweichung vom Lagemass

Es gilt, die Nachteile des letzten Streuungsmasses zu beseitigen.

- Erstens wird die Summe – egal, was genau addiert wird – am Schluss durch n, die Anzahl der Werte, geteilt.
- Zweitens verhindert man, dass sich positive und negative Abweichungen aufheben, indem man das negative Vorzeichen weglässt. Mathematisch heisst das, man verwendet als Abweichung nicht $x_i - \overline{x}$, sondern den Absolutwert oder Betrag davon: $|x_i - \overline{x}|$.

Die Auswirkung:

Note x_i	H_i	$\lvert x_i - \overline{x} \rvert$	$H_i \cdot \lvert x_i - \overline{x} \rvert$
6	2	1.8	3.6
5.5	3	1.3	3.9
5	4	0.8	3.2
4.5	5	0.3	1.5
4	3	0.2	0.6
3.5	2	0.7	1.4
3	3	1.2	3.6
2.5	2	1.7	3.4
2	0	2.2	0
1.5	0	2.7	0
1	1	3.2	3.2
Summe der absoluten Abweichungen			24.4
Mittlere absolute Abweichung (obige Summe durch n=25)			0.976

Für Klasse B erhält man eine mittlere absolute Abweichung von $\frac{9}{16} = 0.5625$, für Klasse C dagegen $\frac{18.8}{20} = 0.94$.

Die so berechnete Streuung ist bei Klasse A also etwas grösser als bei Klasse C.

Die **mittlere absolute Abweichung** ist

$$\frac{H_1 \cdot |x_1 - \overline{x}| + H_2 \cdot |x_2 - \overline{x}| + \ldots + H_k \cdot |x_k - \overline{x}|}{n} \qquad \textbf{[4.2]}$$

und

$$h_1 \cdot |x_1 - \overline{x}| + h_2 \cdot |x_2 - \overline{x}| + \ldots + h_k \cdot |x_k - \overline{x}|,$$

falls die relativen Häufigkeiten h_i bekannt sind.

Dieses Streuungsmass wird durchaus ab und zu verwendet. Aber auch es leidet an einem Nachteil: Das Rechnen mit Beträgen ist eine mühselige Angelegenheit. Wer schon Gleichungen oder sogar Ungleichungen mit Beträgen lösen musste, weiss davon ein Lied zu singen.

4.2.4 Die mittlere quadrierte Abweichung: die theoretische Varianz

Eine weitere Möglichkeit zu verhindern, dass sich Abweichungen nach unten und oben schliesslich aufheben, besteht darin, die Abweichungen zu quadrieren. Minus mal Minus gibt ja auch Plus. Ansonsten verfährt man genau gleich wie bei der mittleren absoluten Abweichung. Für Klasse A sieht das so aus:

Note x_i	H_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$H_i \cdot (x_i - \overline{x})^2$
6	2	1.8	3.24	6.48
5.5	3	1.3	1.69	5.07
5	4	0.8	0.64	2.56
4.5	5	0.3	0.09	0.45
4	3	–0.2	0.04	0.12
3.5	2	–0.7	0.49	0.98
3	3	–1.2	1.44	4.32
2.5	2	–1.7	2.89	5.78
2	0	–2.2	4.84	0.00
1.5	0	–2.7	7.29	0.00
1	1	–3.2	10.24	10.24
Summe der quadrierten Abweichungen				36.00
Mittlere quadrierte Abweichung (obige Summe durch n=25)				1.44

Für Klasse B erhält man $\frac{6.5}{16} = 0.40625$, für Klasse C $\frac{28.8}{20} = 1.44$. Jetzt ist die Streuung bei den Klassen A und C gleich gross!

Dieses Streuungsmass ist von grosser Bedeutung. Es wird als (wahrscheinlichkeits-)theoretische Varianz σ^2 bezeichnet.

Die **(wahrscheinlichkeits-)theoretische Varianz σ^2** ist festgelegt durch

$$\sigma^2 = \frac{H_1 \cdot (x_1 - \overline{x})^2 + H_2 \cdot (x_2 - \overline{x})^2 + \ldots + H_k \cdot (x_k - \overline{x})^2}{n} \qquad \textbf{[4.3]}$$

Wenn die relativen Häufigkeiten h_i bekannt sind, so gilt

$$\sigma^2 = h_1 \cdot (x_1 - \overline{x})^2 + h_2 \cdot (x_2 - \overline{x})^2 + \ldots + h_k \cdot (x_k - \overline{x})^2$$

Aber weshalb wird dieses Streuungsmass mit σ^2 bezeichnet und nicht einfach mit σ?

4.2.5 Die (wahrscheinlichkeits-)theoretische Standardabweichung

Es gibt eben noch ein anderes Streuungsmass, die (wahrscheinlichkeits-)theoretische Standardabweichung, und diese wird mit σ bezeichnet.

Die **(wahrscheinlichkeits-)theoretische Standardabweichung σ** ist festgelegt durch

$$\sigma = \sqrt{\frac{H_1 \cdot (x_1 - \overline{x})^2 + H_2 \cdot (x_2 - \overline{x})^2 + \ldots + H_k \cdot (x_k - \overline{x})^2}{n}} \qquad \textbf{[4.4]}$$

bzw.

$$\sigma = \sqrt{h_1 \cdot (x_1 - \overline{x})^2 + h_2 \cdot (x_2 - \overline{x})^2 + \ldots + h_k \cdot (x_k - \overline{x})^2}\ .$$

Die Standardabweichung ist etwas aufwendiger zu berechnen als die Varianz, weil ja noch die Quadratwurzel gezogen werden muss. Dafür hat die Standardabweichung die-