Glossar: Varianz

Varianz [Stochastik]

Die Varianz ist ein Streuungsmaß, d.h., sie gibt an, wie stark die einzelnen Werte einer Grundgesamtheit um den Mittelwert (Durchschnitt) herum „verstreut liegen".
Stimmen die Werte sehr stark miteinander überein (liegen sie also nahe beieinander), so ist die Streuung gering, weichen sie dagegen erheblich voneinander ab, so ist die Streuung stark.

Die Varianz ist die mittlere quadratische Abweichung der Werte vom arithmetischen Mittel $\bar{x}$
(bzw. vom Erwartungswert μ).

Die Varianz ist neben der Standardabweichung das gebräuchlichste Streuungsmaß.

Bezeichnung: $V(X) = \sigma^2$ (sprich: „Sigma Quadrat")
(im Gegensatz zur Standardabweichung σ)

Die Formel für die Varianz (der Grundgesamtheit) lautet:

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

Bsp.: Wie stark streuen die folgenden Ergebnisse?
(2 ; 6 ; 3)

Rechnung: $\bar{x} = \frac{2+6+3}{3} = 3\frac{2}{3}$

$\sigma^2 = \frac{1}{3}\left((3\frac{2}{3} - 2)^2 + (3\frac{2}{3} - 6)^2 + (3\frac{2}{3} - 3)^2\right) \approx \underline{\underline{2,888889}}$

Die Standardabweichung ist dann
$\sigma = \sqrt{2,888889} \approx 1,69967317$

Rechnung in Tabellenform

i	x_i	$(\bar{x} - x_i)^2$
1	2	2,77777778
2	6	5,44444444
3	3	0,44444444
a.M.	$\bar{x}$= 3,66667	σ ^{2} ≈ 2,88888889

σ ≈ 1,69967317

Varianz einer Zufallsvariable:
Gegeben ist die Zufallsvariable X mit der Ergebnismenge
$\Omega = \{a_1 \,;\, \ldots;\, a_n\}$

und dem Erwartungswert $\mu = \sum_{i=1}^{n} P(X = a_i) \cdot a_i$

Dann ist die Varianz von X

$$\sigma^2 = \sum_{i=1}^{n} P(X = a_i) \cdot (a_i - \mu)^2$$

Was das bedeutet wird wohl eher an einem Beispiel klar:

Bsp: Werfen eines fairen Würfels; Bei einer Sechs gewinnt man 10 € (a_1=10), bei einer anderen Augenzahl über 3 (also 4 oder 5) gewinnt man 1 € (a_2=1), bei einer Augenzahl unter 4 verliert man 5 € (a_3=-5).
$\Omega = \{-4, 1, 10\}$;

$$\mu = \frac{1}{6} \cdot 10 + \frac{1}{3} \cdot 1 + \frac{1}{2} \cdot (-5) = -\frac{1}{2}$$

a_i	$P(X=a_i)$	$(a_i - \mu)^2$	$P(X=a_i) \cdot (a_i - \mu)^2$
10	$\frac{1}{6}$	90,25	15,041667
1	$\frac{1}{3}$	0,25	0,083333
-5	$\frac{1}{2}$	30,25	15,125
			σ ^{2} = 30,25
			σ = $\sqrt{30{,}25}$= 5,5

Bem.: Erheblich einfacher wird es, wenn man es mit einer Binomialverteilung zu tun hat:

Hier gilt die einfache Formel $\sigma^2 = n \cdot p \cdot (1-p)$
bzw. $\sigma^2 = n \cdot p \cdot q$

Bsp.: Binomialverteilung mit n = 100, p = 0,25:
$\sigma^2 = 100 \cdot 0{,}25 \cdot 0{,}75 = 18{,}75$

Bem.: Wenn man allerdings die Streuung einer Grundgesamtheit mit Hilfe von Daten einer Stichprobe bestimmen will, muss man die empirische Varianz verwenden, also durch $n - 1$ teilen statt durch n („empirische Standardabweichung").

FAQ: *Warum benutzt man statt der mittleren quadratischen Abweichung nicht die mittlere Abweichung als Abweichungsmaß?*

Wann man die Differenzen der einzelnen Werte vom Mittelwert benutzen würde, würden sich wegen der unterschiedlichen Vorzeichen gegenseitig genau ausgleichen und es käme immer Null heraus. (Das nennt man die Ausgleichseigenschaft des arithmetischen Mittels).

Warum benutzt man statt der mittleren quadratischen Abweichung nicht die absolute Abweichung als Abweichungsmaß?

Ein Grund hierfür ist, dass dann auch viele kleine Abweichungen verhältnismäßig stark eingehen würden: hundert Werte, die um 0,1 von $\bar{x}$ abweichen würden zum gleichen Ergebnis führen wie ein Wert, der um 10 von $\bar{x}$ abweicht. Man erwartet aber „viele kleine" Abweichungen – diese sollen das Abweichungsmaß daher nicht sonderlich in die Höhe trieben, während einzelne größere Abweichungen stärker eingehen sollen. Das wird durch die Quadratur der Abweichungen erreicht.

Mehr dazu (härterer Tobak): mathepedia