# Lage- und Streuparameter von Zufallsvariablen

**Arno Fehringer, Juni 2018**

**Quellen :**

**Büchter, A.; Henn, H.-W**. : Elementare Stochastik. Eine Einführung in die Mathematik der Daten und des Zufalls . Mathematik für das LEHRAMT. Hrsg.: Reiss, K.; Scharlau, R.; Sonar, T. ;Weigand H.-G. . Springer, 2.Aufl. 2006

# Die Lage- und Streuparameter einer Zufallsvariablen

Sei $X$ eine Zufallsvariable, welche die Werte $x_1, \ldots, x_k$ annehmen kann .

**Erwartungswert** ( von $X$ ) :

$$\boxed{E(X) := \sum_{i=1}^{k} x_i \; P(X=x_i)}$$

Man schreibt auch $E(X) = \mu_X = \mu$ .

**Varianz** ( von $X$ ) :

$$\boxed{V(X) := \sum_{i=1}^{k} (x_i-\mu)^2 \; P(X=x_i)}$$

**Standardabweichung** ( von $X$ ) :

$$\boxed{S(X) := \sqrt{V(X)} = \sqrt{\sum_{i=1}^{k} (x_i-\mu)^2 \; P(X=x_i)}}$$

Man schreibt auch $S(X) = \sqrt{V(X)} = \sigma_X = \sigma$ . Damit ist $V(X) = \sigma^2$ .

Die Zufallsvariablen $X$ , $Y$ , welche die Werte $x_1, \ldots, x_k$ , $y_1, \ldots, y_l$ annehmen können, heißen **unabhängig** , falls gilt :

$$P(X=x_i, Y=y_j) = P(X=x_i)P(Y=y_j) \quad \text{für alle Paare} \quad (x_i, y_j)$$

## Eigenschaften :

1) $E(aX+b) = aE(X) + b$

2) $E(X+Y) = E(X) + E(Y)$

3) $E(XY) = E(X)\,E(Y)$ , falls $X$ , $Y$ unabhängig .

4) $E(X^2) = \sigma^2 + \mu^2$ , $E(X^2) = V(X) + \mu^2$

5) $V(aX+b) = a^2\,V(X)$

6) $V(X+Y) = V(X) + V(Y) + 2\left[E(XY) - \mu_X\mu_Y\right]$

$V(X+Y) = V(X) + V(Y)$ genau dann , wenn $X$ , $Y$ unabhängig .

**Beweise :**

1)
$$E(aX+b) = \sum_{i=1}^{k} (ax_i+b)\,P(X=x_i)$$
$$E(aX+b) = \sum_{i=1}^{k} ax_i\,P(X=x_i) + \sum_{i=1}^{k} b\,P(X=x_i)$$
$$E(aX+b) = a\sum_{i=1}^{k} x_i\,P(X=x_i) + b\sum_{i=1}^{k} P(X=x_i)$$
$$E(aX+b) = aE(X) + b$$

2)
$$E(X+Y) = \sum_{i=1,j=1}^{k,l} (x_i+y_j)\,P(X=x_i, Y=y_j)$$
$$\begin{aligned} E(X+Y) = {} & (x_1+y_1)\,P(X=x_1, Y=y_1) + \ldots + (x_1+y_l)\,P(X=x_1, Y=y_l) \\ & + \\ & . \\ & . \\ & . \\ & + \\ & (x_k+y_1)\,P(X=x_k, Y=y_1) + \ldots + (x_k+y_l)\,P(X=x_k, Y=y_l) \end{aligned}$$

$$
\begin{aligned}
E(X+Y) = {} & x_1\ [P(X=x_1, Y=y_1)\ +\ldots+\ P(X=x_1, Y=y_l)] \\
& + \\
& . \\
& . \\
& . \\
& + \\
& x_k\ [P(X=x_k, Y=y_1)\ +\ldots+\ P(X=x_k, Y=y_l)] \\
& + \\
& y_1\ [P(X=x_1, Y=y_1)\ +\ldots+\ P(X=x_k, Y=y_1)] \\
& + \\
& . \\
& . \\
& . \\
& + \\
& y_l\ [P(X=x_1, Y=y_l)\ +\ldots+\ P(X=x_k, Y=y_l)]
\end{aligned}
$$

$$
\begin{aligned}
E(X+Y) = {} & x_1\ P(X=x_1) \\
& + \\
& . \\
& . \\
& . \\
& + \\
& x_k\ P(X=x_k) \\
& + \\
& y_1\ P(Y=y_1) \\
& + \\
& . \\
& . \\
& . \\
& + \\
& y_l\ P(Y=y_l)
\end{aligned}
$$

$$
E(X+Y) = \sum_{i=1}^{k} x_i\ P(X=x_i) + \sum_{j=1}^{l} y_j\ P(Y=y_j)
$$

$$
E(X+Y) = E(X) + E(Y)
$$

3) $$E(XY) = \sum_{i=1,\, j=1}^{k,\, l} x_i y_j \; P(X=x_i, Y=y_j)$$

$$\begin{aligned}E(XY) = \; & x_1 y_1 \; P(X=x_1, Y=y_1) \; +\ldots+ \; x_1 y_l \; P(X=x_1, Y=y_l) \\ & + \\ & . \\ & . \\ & . \\ & + \\ & x_k y_1 \; P(X=x_k, Y=y_1) \; +\ldots+ \; x_k y_l \; P(X=x_k, Y=y_l)\end{aligned}$$

$$\begin{aligned}E(XY) = \; & x_1 y_1 \; P(X=x_1)P(Y=y_1) \; +\ldots+ \; x_1 y_l \; P(X=x_1)P(Y=y_l) \\ & + \\ & . \\ & . \\ & . \\ & + \\ & x_k y_1 \; P(X=x_k)P(Y=y_1) \; +\ldots+ \; x_k y_l \; P(X=x_k)P(Y=y_l)\end{aligned}$$

$$\begin{aligned}E(XY) = \; & x_1 P(X=x_1) \; \left[y_1 \; P(Y=y_1) \; +\ldots+ \; y_l \; P(Y=y_l)\right] \\ & + \\ & . \\ & . \\ & . \\ & + \\ & x_k P(X=x_k) \; \left[y_1 \; P(Y=y_1) \; +\ldots+ \; y_l \; P(Y=y_l)\right]\end{aligned}$$

$$\begin{aligned}E(XY) = \; & x_1 P(X=x_1) E(Y) \\ & + \\ & . \\ & . \\ & . \\ & + \\ & x_k P(X=x_k) E(Y)\end{aligned}$$

$$E(XY) = E(X) \; E(Y)$$

4) $$E(X^2) = E\left((X-\mu+\mu)^2\right)$$

$$E(X^2) = E\left((X-\mu+\mu)^2\right)$$

$$E(X^2) = E\left((X-\mu)^2 - 2(X-\mu)\mu + \mu^2\right)$$

$$E(X^2) = E\left((X-\mu)^2\right) - E\left(2(X-\mu)\mu\right) + E\left(\mu^2\right)$$

$$E(X^2) = E\left((X-\mu)^2\right) - 2\mu E(X-\mu) + E\left(\mu^2\right)$$

$$E(X^2) = E\left((X-\mu)^2\right) - 2\mu\left(E(X)-E(\mu)\right) + E\left(\mu^2\right)$$

$$E(X^2) = V(X) - 2\mu(\mu-\mu) + \mu^2$$

$$E(X^2) = \sigma^2 + \mu^2 \quad , \quad E(X^2) = V(X) + \mu^2$$

5) Wegen 1) ist $E(aX+b) = aE(X) + b = a\mu+b$ , und es folgt

$$V(aX+b) = \sum_{i=1}^{k} \left(ax_i+b - (a\mu+b)\right)^2 P(X=x_i)$$

$$V(aX+b) = \sum_{i=1}^{k} \left(ax_i - a\mu\right)^2 P(X=x_i)$$

$$V(aX+b) = a^2\sum_{i=1}^{k} \left(x_i - \mu\right)^2 P(X=x_i)$$

$$V(aX+b) = a^2 \, V(X)$$

6) Wegen 4) ist $E(X^2) = \sigma^2 + \mu^2$ , also $\sigma^2 = E(X^2) - \mu^2$ und damit $V(X) = E(X^2) - (E(X))^2$ .

Wendet man dies auf die Variable $X+Y$ an, so folgt

$$V(X+Y) = E\left((X+Y)^2\right) - \left(E(X+Y)\right)^2 \quad .$$

Wegen 2) ist $E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$ , und es folgt

$$V(X+Y) = E\left((X+Y)^2\right) - (\mu_X+\mu_Y)^2$$

$$V(X+Y) = E\left(X^2+2XY+Y^2\right) - \mu_X^2 - 2\mu_X\mu_Y - \mu_Y^2$$

$$V(X+Y) = E(X^2) + 2E(XY) + E(Y^2) - \mu_X^2 - 2\mu_X\mu_Y - \mu_Y^2$$

$$V(X+Y) = E(X^2) - \mu_X^2 + E(Y^2) - \mu_Y^2 + 2\left[E(XY) - \mu_X\mu_Y\right]$$

$$V(X+Y) = V(X) + V(Y) + 2\left[E(XY) - \mu_X\mu_Y\right] \quad .$$

Genau dann, wenn $X$ , $Y$ unabhängig sind gilt nach 3)

$$E(X\,Y) = E(X)\;E(Y) = \mu_X\mu_Y$$

und

$$V(X+Y) = V(X) + V(Y) \quad .$$

# Schätzfunktionen für die Parameter

Gegeben sei eine Zufallsvariable $X$ mit dem Erwartungswert $\mu_X$ und der Varianz $\sigma_X$ und eine zugehörige Grundgesamtheit .

Die Entnahme einer Stichprobe vom Umfang $n$ aus der Grundgesamtheit, kann man sich vorstellen als Realisierung von $n$ **unabhängigen** Zufallsvariablen $X_1 , \dots , X_n$ , die alle identisch wie $X$ verteilt sind .

Das Stichprobenmittel $\overline{X} = \frac{1}{n}\left(X_1 + \dots + X_n\right)$ ist nun ebenfalls eine Zufallsvariable für die man den Erwartungswert und Varianz berechnen kann :

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\left(X_1 + \dots + X_n\right)\right)$$

$$E\left(\overline{X}\right) = \frac{1}{n}E\left(X_1 + \dots + X_n\right)$$

$$E\left(\overline{X}\right) = \frac{1}{n}\left(E\left(X_1\right) + \dots + E\left(X_n\right)\right)$$

$$E\left(\overline{X}\right) = \frac{1}{n}\left(\mu_X + \dots + \mu_X\right)$$

$$E\left(\overline{X}\right) = \frac{1}{n}\ n\mu_X$$

$$\boxed{E\left(\overline{X}\right) = \mu_X} \quad , \quad \mu_{\overline{X}} = \mu_X$$

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\left(X_1 + \dots + X_n\right)\right)$$

$$V\left(\overline{X}\right) = \frac{1}{n^2}V\left(\left(X_1 + \dots + X_n\right)\right)$$

$$V\left(\overline{X}\right) = \frac{1}{n^2}\left(V\left(X_1\right) + \dots + V\left(X_n\right)\right)$$

$$V\left(\overline{X}\right) = \frac{1}{n^2}\left(V\left(X\right) + \dots + V\left(X\right)\right)$$

$$V\left(\overline{X}\right) = \frac{1}{n^2}\ nV\left(X\right)$$

$$\boxed{V\left(\overline{X}\right) = \frac{1}{n}\ V\left(X\right)} \quad , \quad \sigma_{\overline{X}}^{\ 2} = \frac{1}{n}\sigma_X^{\ 2} \quad , \quad \sigma_{\overline{X}} = \frac{1}{\sqrt{n}}\sigma_X$$

Die Varianzen unterscheiden sich, man sagt, die Varianz $V\left(\overline{X}\right)$ des Stichprobenmittels ist **keine erwartungstreue Schätzfunktion** für die Varianz $V\left(X\right)$ .

## Auffinden einer erwartungstreuen Schätzfunktion für die Varianz $V(X)$, die sogenannte empirische Varianz

Betrachte die Funktion $S^2 := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ .

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^{\,2}\right)$$
$$S^2 = \frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - 2\overline{X}\sum_{i=1}^{n}X_i + n\overline{X}^{\,2}\right)$$
$$S^2 = \frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - 2\overline{X}n\overline{X} + n\overline{X}^{\,2}\right)$$
$$S^2 = \frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - 2n\overline{X}^{\,2} + n\overline{X}^{\,2}\right)$$
$$S^2 = \frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^{\,2}\right)$$
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^{\,2}$$

$$E\left(S^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^{\,2}\right)$$
$$E\left(S^2\right) = \frac{1}{n}\sum_{i=1}^{n}E\left(X_i^2\right) - E\left(\overline{X}^{\,2}\right)$$
$$E\left(S^2\right) = \frac{1}{n}\sum_{i=1}^{n}E\left(X^2\right) - E\left(\overline{X}^{\,2}\right)$$
$$E\left(S^2\right) = \frac{1}{n}nE\left(X^2\right) - E\left(\overline{X}^{\,2}\right)$$
$$E\left(S^2\right) = E\left(X^2\right) - E\left(\overline{X}^{\,2}\right)$$

Wegen 4) und $\mu_{\overline{X}} = \mu_X$ ist

$E\left(X^2\right) = V(X) + \mu_X^{\,2}$ , $E\left(\overline{X}^{\,2}\right) = V\left(\overline{X}\right) + \mu_{\overline{X}}^{\,2} = V\left(\overline{X}\right) + \mu_X^{\,2}$ , und es folgt

$$E\left(S^2\right) = V(X) + \mu_X^{\,2} - \left(V\left(\overline{X}\right) + \mu_X^{\,2}\right)$$
$$E\left(S^2\right) = V(X) - V\left(\overline{X}\right)$$

Wegen $V\left(\overline{X}\right) = \frac{1}{n}\,V(X)$ folgt

$$E\left(S^2\right) = V(X) - \frac{1}{n}\,V(X)$$
$$E\left(S^2\right) = \frac{n-1}{n}\,V(X)$$

$$\frac{n}{n-1}\ E\left(S^2\right) = V(X)$$

$$E\left(\frac{n}{n-1}\ S^2\right) = V(X)$$

Man kann also die Funktion $\hat{S}^2 := \dfrac{n}{n-1}\ S^2$ als **erwartungstreue Schätzfunktion** für die Varianz $V(X)$ nehmen .

Es ist

$$\hat{S}^2 = \frac{n}{n-1}\ \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

$$\hat{S}^2 = \frac{1}{n-1}\ \sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

Die Funktion $\hat{S}^2 = \dfrac{1}{n-1}\ \sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ heißt **empirische Varianz** !