

## Kleiner Exkurs in die Mathematik

### **Erwartungswert einer diskreten Zufallsvariable:**

Der Wertevorrat  $W = \{x_1, x_2, x_3, \dots, x_m\}$  der diskreten Zufallsvariable  $X$  sei endlich. Das zugehörige Zufallsexperiment werde  $n$ -mal unabhängig durchgeführt, wobei die  $n$  Realisierungen der Zufallsvariablen  $X$  zu einer Stichprobe vom Umfang  $n$  zusammengefaßt werden. Diese Stichprobe kann dann in einer Häufigkeitstabelle für die  $m$  Merkmalsausprägungen  $x_1, x_2, \dots, x_m$  dargestellt werden mit den relativen Häufigkeiten  $r_n(x_i)$ . Diese Stichprobe besitzt den Mittelwert

$$\bar{x} = \sum_{i=1}^m x_i \cdot r_n(x_i)$$

Für große  $n$  gilt wegen  $r_n(x_i) \approx P(X=x_i)$  für alle  $i$

$$\bar{x} \approx \sum_{i=1}^m x_i \cdot P(X = x_i)$$

Die rechte Seite hängt nur mehr von der Verteilung der Zufallsvariable  $X$  ab, nicht mehr von (der Größe) der Stichprobe. Diese Summe heißt Erwartungswert mit der exakten Definition:

#### **Erwartungswert der diskreten Zufallsvariable $X$ :**

Die diskrete Zufallsvariable  $X$  besitze die Verteilung  $(x_i, P(X=x_i))$ ,  $i = 1, 2, \dots$ ; Dann heißt:

$$E(X) = \mu = \sum x_i \cdot P(X = x_i)$$

der Erwartungswert von  $X$ , falls  $\sum |x_i| \cdot P(X = x_i) < \infty$  ist.

Man sieht daher, daß der Erwartungswert  $E(X)$  einer Zufallsvariablen gleich dem Mittelwert der zugrunde liegenden Verteilung ist. Mit steigendem  $n$  konvergiert der Mittelwert gegen den Erwartungswert.

$$\lim_{n \rightarrow \infty} (\bar{x}) = \mu$$

Für die Praxis bedeutet dieser Umstand, daß die Bestimmung des Mittelwertes um so „genauer“ wird, je größer unsere Stichprobe ist. Um die Genauigkeit bzw. die Ungenauigkeit des experimentell bestimmten Mittelwertes zu kennen, muß man vorerst noch einen weiteren Parameter betrachten, nämlich die Varianz  $V(X)$ , bzw. die Standardabweichung  $\sigma(x)$ .

Die Varianz selber ist wieder ein „Erwartungswert“ und zwar der Erwartungswert der Quadrate der Abweichung vom Mittelwert, in genauer mathematischer Definition geschrieben als:

Die **Varianz** einer Zufallsvariable  $X$  ist der Erwartungswert von  $(X-\mu)^2$ ; üblicherweise geschrieben als  $V(X)$ . Die Standardabweichung  $\sigma = \sqrt{V(X)}$ .

$$\begin{aligned} V(X) &= E[(X - \mu)^2] \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

in anderer Schreibweise:

$$\begin{aligned}
 V(X) &= E[(X - \mu)^2] \\
 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2
 \end{aligned}$$

$$\sigma = \sqrt{V(X)}$$

**Achtung:** Für die Berechnung der Varianz und der Standardabweichung nach diesem Verfahren ist die Kenntnis des Erwartungswertes der Verteilung notwendig. Normalerweise ist das in experimentellen Untersuchungen nicht der Fall, die Berechnung der Varianz und der Standardabweichung ist daher zu modifizieren!

Experimentell ermittelte Varianz und Standardabweichung

$$V(X) = S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Für die Berechnung der experimentell ermittelten Standardabweichung steht im Nenner n-1 im Vergleich zu n für die Standardabweichung der zugrunde liegenden Verteilung. Warum ist das so ?

Betrachten wir zunächst die Berechnung der Summe der Abweichungsquadrate vom Mittelwert.

$$\begin{aligned}
 \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2 \\
 &= \sum_{i=1}^n [(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2 \cdot (X_i - \mu)(\bar{X} - \mu)] \\
 &= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2 \cdot (\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) \\
 &= \sum_{i=1}^n (X_i - \mu)^2 + n \cdot (\bar{X} - \mu)^2 - 2n \cdot (\bar{X} - \mu)^2 \\
 &= \sum_{i=1}^n (X_i - \mu)^2 - n \cdot (\bar{X} - \mu)^2
 \end{aligned}$$

Unter Anwendung der Erwartungswerte auf beiden Seiten der Identität

$$\begin{aligned}
 E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= E\left[\sum_{i=1}^n (X_i - \mu)^2\right] - n \cdot E[(\bar{X} - \mu)^2] \\
 &= \sum_{i=1}^n E[(X_i - \mu)^2] - n \cdot V(\bar{X}), \text{ weil } E(\bar{X}) = \mu. \\
 &= \sum_{i=1}^n V(X_i) - n \cdot \frac{\sigma^2}{n} \\
 &= n \cdot \sigma^2 - \sigma^2 \\
 &= (n-1) \cdot \sigma^2
 \end{aligned}$$

Durch Division beider Seiten durch (n-1):

$$E \left[ \frac{\sum_{i=1}^n (X - \bar{X})^2}{n-1} \right] = \sigma^2, \quad n > 1$$

erhalten wir die Standardabweichung der Verteilung.

Die **Varianz einer Stichprobe** mit den Realisierungen  $x_1, x_2, x_3, \dots, x_n$  der Zufallsvariable  $X$  wird mit  $S^2$  bezeichnet und definiert als:

$$s^2 = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n-1}, \quad \text{bzw. } E(s^2) = \sigma^2 = V(X)$$

Der Erwartungswert der experimentell ermittelten Varianz ist also die Varianz der Verteilung, ein nicht uninteressantes, vor allem aber nicht unwichtiges Ergebnis.

### Varianz und Standardabweichung des Mittelwertes $\bar{X}$

Alle  $X: X_1, X_2, \dots, X_n$  sind unabhängige Zufallsvariablen mit der gleichen Varianz  $\sigma^2$ , dann ist

$$\begin{aligned} V(X_1 + X_2 + \dots + X_n) &= V(X_1) + V(X_2) + \dots + V(X_n) \\ &= n \cdot \sigma^2 \end{aligned}$$

Wenn  $U$  eine Zufallsvariable und  $a$  eine Konstante ist, dann ist  $V(a \cdot U) = a^2 \cdot V(U)$ .

Mit  $U = X_1 + X_2 + \dots + X_n$  und  $a = 1/n$ , ergeben diese Regeln:

$$\begin{aligned} V(\bar{X}) &= \frac{V(X_1 + X_2 + \dots + X_n)}{n^2} \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Die Varianz des Mittelwertes ist also genau um den Faktor  $1/n$  kleiner als die Varianz der zugrunde liegenden Verteilung. Die Standardabweichung des Mittelwertes  $\bar{X}$  ist daher um den Faktor  $1/\sqrt{n}$  kleiner als die Standardabweichung von  $X$ .

### Beispiel und Fragen

1. Was ist der Erwartungswert beim Werfen eines Würfels mit den Zahlen 1, ..., 6.
2. Was ist die Standardabweichung des Mittelwertes einer Probe mit  $n=20$ ,  $n = 50$ . Stellen Sie den funktionellen Zusammenhang  $s = f(n)$  graphisch dar.