# Statistische Maßzahlen

Steffen Solyga*

22. Februar 2005

## 1 Definitionen

### 1.1 Arithmetisches Mittel

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad n = 1, 2, \ldots \tag{1}$$

### 1.2 Empirische Varianz

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x}_n)^2, \qquad n = 2, 3, \ldots \tag{2}$$

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\,\overline{x}_n^2, \qquad n = 2, 3, \ldots \tag{3}$$

### 1.3 Empirische Standardabweichung

$$s_n = \sqrt{s_n^2}, \qquad n = 2, 3, \ldots \tag{4}$$

## 2 Rekursionsformeln

### 2.1 Arithmetisches Mittel

$$\overline{x}_{n+1} = \frac{n}{n+1}\,\overline{x}_n + \frac{x_{n+1}}{n+1}, \qquad n = 1, 2, \ldots \tag{5}$$

### 2.2 Empirische Varianz

$$s_{n+1}^2 = \frac{n-1}{n}\,s_n^2 + \overline{x}_n^2 - \frac{n+1}{n}\,\overline{x}_{n+1}^2 + \frac{x_{n+1}^2}{n}, \qquad n = 2, 3, \ldots \tag{6}$$

*solyga@absinth.net

## 3 Alternative Varianz

Der gemäß (2) definierten empirischen Varianz haftet ein entscheidender Makel an: Sie existiert erst, sobald wenigstens 2 Werte $x_i$ vorliegen. Spielt ihre Erwartungstreue keine Rolle, definiert man sie sinnvoller als[1]

$$s_n^2 \quad = \quad \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2, \qquad n = 1, 2, \ldots . \tag{7}$$

Wegen (1) gilt

$$\sum_{i=1}^{n} x_i \quad = \quad n\overline{x}_n, \qquad n = 1, 2, \ldots , \tag{8}$$

und aus (7) folgt

$$\begin{aligned}
ns_n^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x}_n + \overline{x}_n^2) && (9)\\
&= \sum_{i=1}^{n} x_i^2 - 2\overline{x}_n\sum_{i=1}^{n} x_i + \overline{x}_n^2\sum_{i=1}^{n} 1 && (10)\\
&= \sum_{i=1}^{n} x_i^2 - 2\overline{x}_n n\overline{x}_n + \overline{x}_n^2 n && (11)\\
&= \sum_{i=1}^{n} x_i^2 \; - \; n\overline{x}_n^2 && (12)
\end{aligned}$$

und mithin

$$s_n^2 \quad = \quad \frac{1}{n}\sum_{i=1}^{n} x_i^2 \; - \; \overline{x}_n^2. \tag{13}$$

Hier ist lediglich über die $x_i^2$, nicht jedoch über die $\overline{x}_n^2$ zu summieren! Analog zu (8) gilt

$$\sum_{i=1}^{n} x_i^2 \quad = \quad n(s_n^2 + \overline{x}_n^2), \qquad n = 1, 2, \ldots . \tag{14}$$

Für $n := n + 1$ folgt aus (7) bzw. (13)

$$\begin{aligned}
s_{n+1}^2 &= \frac{1}{n+1}\sum_{i=1}^{n+1}(x_i - \overline{x}_{n+1})^2, && (15)\\
&= \frac{1}{n+1}\sum_{i=1}^{n+1} x_i^2 \; - \; \overline{x}_{n+1}^2, && (16)\\
&= \frac{1}{n+1}\sum_{i=1}^{n} x_i^2 + \frac{x_{n+1}^2}{n+1} \; - \; \overline{x}_{n+1}^2, && (17)\\
&= \frac{n(s_n^2 + \overline{x}_n^2)}{n+1} + \frac{x_{n+1}^2}{n+1} \; - \; \overline{x}_{n+1}^2. && (18)
\end{aligned}$$

[1] Wegen $(n + 1)/n \to 1$ stimmt (7) mit (2) für große $n$ überein.

## 4 Alternative Rekursionsformeln

Definiert man

$$\bar{x}_0 := 0, \tag{19}$$

$$s_0^2 := 0, \tag{20}$$

so erhält man folgende nicht-singuläre Darstellung der statistischen Maßzahlen:

$$\bar{x}_{n+1} = \frac{n\bar{x}_n + x_{n+1}}{n+1}, \qquad n = 0, 1, \ldots, \tag{21}$$

$$s_{n+1}^2 = \frac{n(s_n^2 + \bar{x}_n^2) + x_{n+1}^2}{n+1} - \bar{x}_{n+1}^2, \qquad n = 0, 1, \ldots. \tag{22}$$

Diese eignet sich vorzüglich für die fortlaufende Berechnung: Die Initialisierung beim Programmstart erfolgt gemäß (19) und (20), und mit dem Verfügbarsein des ersten Wertes ($x_1$) läßt sich zunächst das arithmetische Mittel gemäß (21) und anschließend die empirische Varianz gemäß (22) berechnen usw..

## Literatur


[1] I. N. Bronstein, K. A. Semendjajew: Taschenbuch der Mathematik. BSB B. G. Teubner Verlagsgesellschaft, 22. Auflage, Leipzig 1985

[2] Beyer, Hackel, Pieper, Tiedge: Wahrscheinlichkeitsrechnung und mathematische Statistik. Mathematik für Ingenieure, Naturwissenschaftler, Ökonomen und Landwirte, Band 17. BSB B. G. Teubner Verlagsgesellschaft, 3. Auflage, Leipzig 1982