Mittelwert und Standardabweichung

Wilfried Mann, Mettmann

Der Mittelwert ist ein statistisches Merkmal, der die allgemeine Niveaulage in der betrachteten Gesamtheit (Grundgesamtheit oder Stichprobe) charakterisiert. Die Standardabweichung ist ein Parameter zu Beurteilung von arithmetischen Mittelwerten.

1 Arten von Mittelwerten

Es gibt eine Vielzahl von Mittelwerten, die in der Statistik ganz unterschiedlich verwendet werden. Neben den vier hier aufgeführten wichtigen Mitteln gehören auch z.B. geometrische, harmonische, gleitende und logarithmische Mittel dazu.

Arithmetisches Mittel

Für metrische Daten ist der arithmetische Mittelwert ($\bar{x}_{arithm}$), als Durchschnitt, der gebräuchlichste Mittelwert. Dieser besitzt für die Statistik auch wünschenswerte schätztheoretische Eigenschaften wie z.B. Erwartungstreue. Sind n Ausprägungen x_i (i = 1, …, n) eines metrischen Merkmals gegeben, so ist das arithmetische Mittel definiert durch

$$\bar{x}_{arithm} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Das arithmetische Mittel ist also gleich dem Gesamtmerkmalsbetrag dividiert durch die Anzahl der Merkmalsträger.

Median

Der Median oder Zentralwert ($\bar{x}_{med}$) ist ein Lageparameter und stellt den mittleren Wert einer der Größe nach geordneten Zahlenreihe dar. Die Formel hängt davon ab, ob es sich um eine gerade oder ungerade Anzahl von Kaufpreisen handelt. Der einfachere Fall ist eine ungerade Anzahl, es gilt:

$$\bar{x}_{med} = x_{\left(\frac{n+1}{2}\right)}$$

Bei einer geraden Anzahl wird einfach das arithmetische Mittel der beiden in der Mitte liegenden Werte gebildet. Die Formel lautet:

$$\bar{x}_{med} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$$

Modus

Der Modus oder Modalwert ist die Ausprägung mit höchster Häufigkeit. Er ist nur für nicht-stetige Daten definiert. Bei klassierten Variablentypen kann nur die Gruppe mit der größten Häufigkeit angegeben werden.

Gewichtetes (gewogenes) arithmetisches Mittel

Die einzelnen Merkmalswerte werden mit Gewichten versehen. Bei Gewichten, z.B. als absolute Häufigkeiten (*H*), ergibt sich:

$$\bar{x}_{gew} = \frac{1}{n}\sum_{i=1}^{m} x_i H_i = \frac{x_1 H_1 + x_2 H_2 + \cdots + x_m H_m}{n} \quad \text{mit} \quad n = \sum_{i=1}^{m} H_i$$

2 Standardabweichung

Die Standardabweichung ist ein Streuungsmaß. Sie hat dieselbe Dimension wie die Merkmalswerte selbst und errechnet sich als positive Wurzel der Varianz.

Varianz

Maßzahl zur Charakterisierung der Streuung einer theoretischen oder empirischen Verteilung. Hierbei wird ermittelt, wie weit die Daten (x_1, x_2, …, x_n) vom Mittelwert abweichen. Ist der »wahre« Mittelwert oder Schätzwert (µ) bekannt, errechnet sich die Varianz (σ^2) zu:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Meistens wird die Varianz jedoch als Stichprobenvarianz (s^2) aus dem arithmetischen Mittelwert ($\bar{x}$) abgeleitet:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Die **Standardabweichung** ist dann jeweils

$\sigma = \sqrt{\sigma^2}$ bzw. als empirische Standardabweichung: $s = \sqrt{s^2}$

Variationskoeffizient

Auch mit relativer Abweichung (v) bezeichnet, ist er das Verhältnis zwischen Standardabweichung und Mittelwert und dimensionslos.

$v = \frac{s}{\bar{x}}$ Er kann auch in Prozent angegeben werden: $cv = \frac{100s}{\bar{x}}$

Standardabweichung des Mittels

Diese Größe ist ein Maß für die Genauigkeit des Mittelwertes.

$$s_{mittel} = \frac{s}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

In der Regel ist s_{mittel} kleiner als s. Höhere Fallzahlen führen immer zu höheren Genauigkeiten.

3 Anwendung

Mittelwerte werden zur Bestimmung der Niveaulage von Kaufpreisen und deren beschreibende Merkmale wie Grundstücksfläche, Wohnfläche, Baujahr usw. genutzt.
Stehen viele Kauffälle zur Verfügung, wird der arithmetische Mittelwert ($\bar{x}$) berechnet. Zur Abschätzung der Qualität des Datenmaterials dient dann die Standardabweichung (s).
Ausreißer werden häufig bei > +/- 3-s definiert. Dies kann allerdings nicht automatisch erfolgen, da in der Praxis die Grenzen, bedingt durch marktgängiges Verhalten, auch deutlich darüber liegen können. Deshalb sollten Ausreißer erst nach Anpassung der Kaufpreise (z.B. bei Vergleichspreisen) sachverständig ausgeschlossen werden.
Einen Qualitätsvergleich zwischen verschiedenen Wertrelationen (Kaufpreise, Liegenschaftszinssätze usw.) ermöglicht elegant der dimensionslose Variationskoeffizient (cv).
Liegen für einen Preisvergleich nur wenige Kauffälle vor (n < 10), kann der Median bei der Vergleichspreisfindung herangezogen werden. Auch das gewichtete Mittel ist nach der Vergleichswert-Richtlinie (VW-RL) ein geeigneter Mittelwert.
Der Modalwert kommt bei gruppierten Variablen wie Wohnlagen und Ausstattungsklassen zur Anwendung.