WiSt-Taschenbücher

Statistische Formeln und Tabellen

Kompakt für Wirtschaftswissenschaftler

von
Prof. Dr. Josef Bleymüller, Prof. Dr. Rafael Weißbach

13., überarbeitete Auflage

Verlag Franz Vahlen München 2015

Verlag Franz Vahlen im Internet:
www.vahlen.de
ISBN 978 3 8006 4962 4

Zu Inhaltsverzeichnis

Arithmetisches Mittel μ

Bei N *Einzelwerten*

$$a_1, a_2, \ldots, a_N$$

ist das arithmetische Mittel definiert als

$$\mu = \frac{1}{N}\sum_{i=1}^{N} a_i \,.$$

Bei einer *Häufigkeitsverteilung* mit k verschiedenen Werten

$$x_1, x_2, \ldots, x_k$$

ergibt sich das (gewogene) arithmetische Mittel zu

$$\mu = \frac{1}{N}\sum_{i=1}^{k} x_i h_i \quad \text{bzw.} \quad \mu = \sum_{i=1}^{k} x_i f_i \,.$$

Bei einer *Häufigkeitsverteilung klassifizierter Daten* ergibt sich mithilfe der Klassenmitten

$$x_1', x_2', \ldots, x_k'$$

näherungsweise

$$\mu = \frac{1}{N}\sum_{i=1}^{k} x_i' h_i \quad \text{bzw.} \quad \mu = \sum_{i=1}^{k} x_i' f_i \,.$$

Für eine Grundgesamtheit, die aus k *Teilgesamtheiten* mit den Umfängen $N_1, N_2, \ldots, N_k$ und den arithmetischen Mitteln $\mu_1, \mu_2, \ldots, \mu_k$ besteht, ergibt sich das arithmetische Mittel zu

$$\mu = \sum_{i=1}^{k} \frac{N_i}{N}\mu_i \quad \text{bzw.} \quad N = \sum_{i=1}^{k} N_i \,.$$

Median Me und Quartile Q_1, Q_2 und Q_3

Zunächst werden die *Einzelwerte* $a_1, a_2, \ldots, a_N$ so umgeordnet, dass gilt

$$a_{[1]} \le a_{[2]} \le \ldots \le a_{[N]} \,.$$

Dann ist bei ungeradem N

$$\mathrm{Me} = a_{\left[\frac{N+1}{2}\right]}$$

und bei geradem N

$$\mathrm{Me} = \frac{1}{2}\left(a_{\left[\frac{N}{2}\right]} + a_{\left[\frac{N}{2}+1\right]}\right) .$$

Für großes N kann als Median der *größte* Merkmalswert $a_{[k]}$ verwendet werden, für den

$$\mathrm{F}(a_{[k]}) \leq 0{,}5$$

gilt, wobei $\mathrm{F}(a_{[k]})$ der Wert der Summenhäufigkeitsfunktion für $a_{[k]}$ ist.

Analog ist das 1. Quartil Q_1 der größte Merkmalswert $a_{[j]}$, für den

$$\mathrm{F}(a_{[j]}) \leq 0{,}25$$

und das 3. Quartil Q_3 der *größte* Merkmalswert $a_{[l]}$, für den

$$\mathrm{F}(a_{[l]}) \leq 0{,}75$$

gilt.

Bei *klassifizierten Daten* ergibt sich der feinberechnete Median aus der Klassenuntergrenze x_i^u und der Klassenobergrenze x_i^o derjenigen Klasse i, in der die Summenhäufigkeitsfunktion den Wert 0,5 erreicht:

$$\mathrm{Me} = x_i^u + \frac{0{,}5 - \mathrm{F}(x_i^u)}{\mathrm{F}(x_i^o) - \mathrm{F}(x_i^u)}(x_i^o - x_i^u) .$$

In analoger Weise ergeben sich die feinberechneten Quartile zu

$$Q_1 = x_k^u + \frac{0{,}25 - \mathrm{F}(x_k^u)}{\mathrm{F}(x_k^o) - \mathrm{F}(x_k^u)}(x_k^o - x_k^u) ,$$

wobei k diejenige Klasse ist, in der die Summenhäufigkeitsfunktion den Wert 0,25 erreicht, und

$$Q_3 = x_l^u + \frac{0{,}75 - \mathrm{F}(x_l^u)}{\mathrm{F}(x_l^o) - \mathrm{F}(x_l^u)}(x_l^o - x_l^u) ,$$

wobei l diejenige Klasse ist, in der die Summenhäufigkeitsfunktion den Wert 0,75 erreicht.

Q_2 entspricht dem Median Me.

Modus Mo

Der Modus Mo ist als die häufigste Merkmalsausprägung definiert. Bei klassifizierten Daten wird als Modus die Klassenmitte der Klasse mit der größten Säulenhöhe im Histogramm gewählt.

Geometrisches Mittel G

Bei *Einzelwerten* ist das geometrische Mittel definiert als

$$\mathrm{G} = \sqrt[N]{a_1 \cdot a_2 \cdot \ldots \cdot a_N} \quad \text{bzw.} \quad \log \mathrm{G} = \frac{1}{N} \sum_{i=1}^{N} \log a_i \,.$$

Für *Häufigkeitsverteilungen* ergibt sich

$$\mathrm{G} = \sqrt[N]{x_1^{h_1} \cdot x_2^{h_2} \cdot \ldots \cdot x_k^{h_k}} \quad \text{bzw.}$$

$$\log \mathrm{G} = \frac{1}{N} \sum_{i=1}^{k} h_i \log x_i = \sum_{i=1}^{k} f_i \log x_i \,.$$

In der folgenden Tabelle wird angegeben, bei welchen Skalenniveaus die Berechnung des entsprechenden Mittelwertes sinnvoll ist.

Mittelwerte	Skala			
	Nominal-skala	Ordinal-skala	Intervall-skala	Verhältnis-skala
Modus	×	×	×	×
Median und Quartile		×	×	×
Arithmetisches Mittel			×	×
Geometrisches Mittel				×

Varianz σ^2 und Standardabweichung σ

Bei *Einzelwerten* ist die Varianz definiert als

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} a_i^2 - \mu^2 .$$

Bei *Häufigkeitsverteilungen* erhält man die Varianz zu

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \mu)^2 h_i \;=\; \frac{1}{N}\sum_{i=1}^{k} x_i^2 h_i - \mu^2$$
$$= \frac{1}{N}\sum_{i=1}^{k} x_i^2 h_i - \left(\frac{\sum_{i=1}^{k} x_i h_i}{N}\right)^2$$

bzw.

$$\sigma^2 = \sum_{i=1}^{k}(x_i - \mu)^2 f_i \;=\; \sum_{i=1}^{k} x_i^2 f_i - \mu^2$$
$$= \sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 .$$

Bei einer *Häufigkeitsverteilung klassifizierter Daten* ergibt sich die Varianz näherungsweise zu

$$\sigma^2 \;=\; \frac{1}{N}\sum_{i=1}^{k}(x_i' - \mu)^2 h_i = \frac{1}{N}\sum_{i=1}^{k} x_i'^2 h_i - \mu^2$$
$$= \sum_{i=1}^{k}(x_i' - \mu)^2 f_i = \sum_{i=1}^{k} x_i'^2 f_i - \mu^2 .$$

Bei einer *unimodalen (eingipfligen) Verteilung* und einer konstanten Klassenbreite Δx führt die Sheppard-Korrektur zum *besseren Näherungswert*

$$\sigma^2_{korr.} = \sigma^2 - \frac{(\Delta x)^2}{12} .$$

Für eine Grundgesamtheit, die aus k *Teilgesamtheiten* mit den Umfängen $N_1, N_2, \ldots, N_k$, den arithmetischen Mitteln $\mu_1, \mu_2, \ldots, \mu_k$ und den Varianzen $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, besteht, ergibt sich die Varianz zu

$$\sigma^2 = \sum_{i=1}^{k}\frac{N_i}{N}\sigma_i^2 + \sum_{i=1}^{k}\frac{N_i}{N}(\mu_i - \mu)^2$$

mit

$$N = \sum_{i=1}^{k} N_i \quad \text{bzw.} \quad \mu = \sum_{i=1}^{k} \frac{N_i}{N} \mu_i \,.$$

Die *Standardabweichung* σ ergibt sich jeweils als

$$\sigma = \sqrt{\sigma^2} \,.$$

Standardisierung

Aus den *Einzelwerten* $a_1, a_2, \ldots, a_N$ werden die *standardisierten Einzelwerte* z_i nach der Formel

$$z_i = \frac{a_i - \mu}{\sigma} \quad (i = 1, \ldots, N)$$

berechnet, wobei

$$\mu = \frac{1}{N} \sum_{i=1}^{N} a_i \quad \text{und} \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (a_i - \mu)^2}$$

ist.

Die *standardisierten Einzelwerte* z_i $(i = 1, \ldots, N)$ besitzen das arithmetische Mittel 0 und die Varianz 1.

Variationskoeffizient VC

$$\mathrm{VC} = \frac{\sigma}{\mu} \quad \text{bzw.} \quad \mathrm{VC} = \frac{\sigma}{\mu} 100\%$$

Mittlere absolute Abweichung MAD bezogen auf μ

Bei *Einzelwerten* ergibt sich

$$\mathrm{MAD} = \frac{1}{N} \sum_{i=1}^{N} |a_i - \mu|$$

und bei einer *Häufigkeitsverteilung*

$$\mathrm{MAD} = \frac{1}{N} \sum_{i=1}^{k} |x_i - \mu| h_i$$

bzw.

$$\mathrm{MAD} = \sum_{i=1}^{k} |x_i - \mu| f_i \,.$$

Spannweite R

Die *Einzelwerte* $a_1, a_2, \ldots, a_N$ werden der Größe nach angeordnet, so dass gilt:

$$a_{[1]} \leq a_{[2]} \leq \ldots \leq a_{[N]} \ .$$

Dann ist

$$R = a_{[N]} - a_{[1]} \ .$$

Quartilsabstand QA

$$\mathrm{QA} = Q_3 - Q_1$$

Mittlerer Quartilsabstand MQA

$$\mathrm{MQA} = \frac{Q_3 - Q_1}{2}$$

Quartilsdispersionskoeffizient QDC

$$\mathrm{QDC} = \frac{Q_3 - Q_1}{Q_3 + Q_1}\, 100\%$$

In der folgenden Tabelle wird angegeben, bei welchen Skalenniveaus eine Berechnung des entsprechenden Streuungsmaßes sinnvoll ist.

Streuungsmaße	Skala			
	Nominal-skala	Ordinal-skala	Intervall-skala	Verhältnis-skala
Spannweite		×	×	×
Quartilsabstand		×	×	×
Mittlerer Quartilsabstand		×	×	×
Mittlere absolute Abweichung			×	×
Varianz, Standardabweichung			×	×
Variationskoeffizient				×
Quartilsdispersionskoeffizient				×

A, B und E bezeichnen Ereignisse; S ist der Ereignisraum.

Klassische Wahrscheinlichkeitsdefinition

Sind alle Elementarereignisse gleichmöglich, so ist

$$\mathrm{W}(A) = \frac{\text{Anzahl der für A günstigen Fälle}}{\text{Anzahl aller gleichmöglichen Fälle}} \, .$$

Statistische Wahrscheinlichkeitsdefinition

$$\mathrm{W}(A) = \lim_{n\to\infty} f_n(A) = \lim_{n\to\infty} \frac{h_n(A)}{n}$$

Axiomatische Wahrscheinlichkeitsdefinition

Axiome von Kolmogorov:

(1) $0 \leq \mathrm{W}(A) \leq 1$ für $A \subset S$

(2) $\mathrm{W}(S) = 1$

(3) $\mathrm{W}(A \cup B) = \mathrm{W}(A) + \mathrm{W}(B)$ für $A \cap B = \emptyset$

Aus Axiom (3) ergibt sich die **Beziehung**

$$\mathrm{W}(A_1 \cup A_2 \cup \ldots \cup A_n) = \mathrm{W}(A_1) + \mathrm{W}(A_2) + \ldots + \mathrm{W}(A_n)$$

für

$$A_i \cap A_j = \emptyset \quad (i \neq j) \, .$$

Gegenwahrscheinlichkeit

Für $\overline{A}$, das Komplementärereignis von A, gilt

$$\mathrm{W}(\overline{A}) = 1 - \mathrm{W}(A) \, .$$

De Morgansche Gesetze

Aus den mengentheoretischen Beziehungen

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \quad \text{und} \quad \overline{A \cap B} = \overline{A} \cup \overline{B}$$

lassen sich folgende Regeln ableiten:

$$\begin{aligned} \mathrm{W}(A \cup B) &= 1 - \mathrm{W}(\overline{A} \cap \overline{B}) \quad \text{und} \\ \mathrm{W}(A \cap B) &= 1 - \mathrm{W}(\overline{A} \cup \overline{B}) \, . \end{aligned}$$

Additionssatz

$$W(A \cup B) = W(A) + W(B) - W(A \cap B)$$

Bedingte Wahrscheinlichkeit

Für $W(A) > 0$ ist die bedingte Wahrscheinlichkeit des Ereignisses B unter der Bedingung A definiert als

$$W(B/A) = \frac{W(A \cap B)}{W(A)} .$$

Stochastische Unabhängigkeit

Zwei Ereignisse A, B heißen stochastisch unabhängig genau dann, wenn

$$W(B/A) = W(B/\overline{A}) \vee W(A/B) = W(A/\overline{B}) ,$$

bzw. $W(A \cap B) = W(A) \cdot W(B)$ gilt.

Multiplikationssatz

Für *stochastisch unabhängige Ereignisse* A, B gilt

$$W(A \cap B) = W(A) \cdot W(B) .$$

Für *stochastisch abhängige Ereignisse* A, B gilt

$$W(A \cap B) = W(A) \cdot W(B/A) = W(B) \cdot W(A/B).$$

Theorem von der totalen Wahrscheinlichkeit

Wenn $A_1 \cup A_2 \cup \ldots \cup A_n = S$ und $A_i \cap A_j = \emptyset$ für $i \neq j$ gilt, dann ist für $E \subset S$

$$W(E) = \sum_{i=1}^{n} W(A_i) \cdot W(E/A_i) .$$