

Median aus Relativen Häufigkeiten - Grundwissen



Gegeben sei eine univariate statistische Erhebung mit

- einer Grundgesamtheit mit dem Erhebungsumfang n ,
- einem quantitativen Merkmal X mit m verschiedenen Merkmalsausprägungen $a_1 ; \dots ; a_m$ und
- den Relativen Häufigkeiten $h(a_1), \dots, h(a_m)$ der einzelnen Merkmalsausprägungen.

Dann bestimmt sich den **Median** (oder **Zentralwert**) x_{med} oder \tilde{x} (lies: ‚x-Schlange‘) (der gewonnenen Daten) als diejenige Zahl, für die die eine Hälfte der Messwerte kleiner oder gleich und die andere Hälfte größer oder gleich ist, d.h. durch die Bedingung

$$\underbrace{x_1 \leq \dots \leq}_{\text{die Hälfte der 'kleineren' Messwerte}} \tilde{x} \leq \underbrace{\dots \leq x_n}_{\text{die Hälfte der 'größeren' Messwerte}}$$

- Ist m ungerade, dann gilt $\tilde{x} = x_{\frac{n+1}{2}}$, \tilde{x} ist hier auf jeden Fall ein Messwert; ist m gerade,

dann setzt man $\tilde{x} = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$, \tilde{x} ist in diesem Fall nicht unbedingt ein Messwert.

Beispiel: Gegeben sind die Relativen Häufigkeiten

a_i	1,3	1,4	1,5	1,6	1,7	1,8	1,9
$h(a_i)$	5%	20%	25%	15%	20%	10%	5%

Berechne den Median \tilde{x} der gewonnenen Daten.

Da 50% der Messwerte kleiner oder gleich 1,5 und 50% der Messwerte größer oder gleich 1,6

sind, ergibt sich $\tilde{x} = \frac{1,5 + 1,6}{2} = \frac{3,1}{2} = 1,55$.