

Für diskrete Variable X mit möglichen Werten a_1, \dots, a_K liege Stichprobe x_1, \dots, x_n vom Umfang n vor.

Absolute Häufigkeiten: $h_k = |\{j \in \{1, \dots, n\} : x_j = a_k\}|, k = 1, \dots, K$.

Relative Häufigkeiten: $r_k = \frac{h_k}{n}, k = 1, \dots, K$.

Tabelle 1: Klassifizierte oder gruppierte Daten (für u.a. Histogramm)

Klassen – Nr.	Klasse	Klassenrepräsentant	Häufigkeit	relative Häufigkeit	Klassenbreite	Klassendichte
1	$[a_0, a_1)$	x_1^*	h_1	r_1	$d_1 = a_1 - a_0$	$\frac{r_1}{d_1}$
2	$[a_1, a_2)$	x_2^*	h_2	r_2	$d_2 = a_2 - a_1$	$\frac{r_2}{d_2}$
\vdots	\vdots	\vdots	\vdots	\vdots		
k	$[a_{k-1}, a_k)$	x_k^*	h_k	r_k	$d_k = a_k - a_{k-1}$	$\frac{r_k}{d_k}$
\vdots	\vdots	\vdots	\vdots	\vdots		
K	$[a_{K-1}, a_K)$	x_K^*	h_K	r_K	$d_K = a_K - a_{K-1}$	$\frac{r_K}{d_K}$

Mittelwert

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{x_1 + \dots + x_n}{n}$$

Median:

Sei $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ die sog. **geordnete Stichprobe**:

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & , n \text{ ungerade} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2} & , n \text{ gerade.} \end{cases}$$

Empirische Quantile

Für jedes p mit $0 < p < 1$ ist das **empirische p*100%-Quantil** x_p der Stichprobe x_1, \dots, x_n der $n \cdot p$ kleinste Wert der Stichprobe. D.h.: Finde zu gegebenem p die natürliche Zahl k mit

$$n \cdot p \leq k < n \cdot p + 1 \quad \text{und setze : } x_p = x_{(k)} \text{ für } p \neq 0,5 \text{ und } x_{0,5} = \tilde{x}.$$

$x_{0,25}$ heißt auch **unteres Quartil** und $x_{0,75}$ heißt **oberes Quartil**.

Mittelwert bei gruppierten Daten

Beachte Tabelle 1. Dann setzt man.

$$\bar{x}_{\text{gruppiert}} = \sum_{i=1}^K x_i^* \cdot r_i.$$

Median bei gruppierten Daten

Beachte Tabelle 1. Der Median der klassifizierten Daten ist der Klassenrepräsentant (Mitte der Klasse) derjenigen Klasse, in der der Median liegt.

5-Punkte-Zusammenfassung:

$$x_{(1)} , x_{0,25} , \tilde{x} , x_{0,75} , x_{(n)} .$$

Die graphische Darstellung der 5-Punkte-Zusammenfassung nennt man **Boxplot**.

Empirische Varianz

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

Empirische Streuung oder Standardabweichung $s_x = \sqrt{s_x^2}$

Variationskoeffizient

$$V_x = \frac{s_x}{\bar{x}}$$

Interquartilsabstand

$$x_{0,75} - x_{0,25} .$$

Quartilskoeffizient der Schiefe:

$$QKS_x := \frac{(x_{0,75} - \tilde{x}) - (\tilde{x} - x_{0,25})}{x_{0,75} - x_{0,25}} .$$