Zusammenfassung Beschreibende Statistik.

Stichprobe:
Wird der Teil einer Gesamtheit befragt, dann spricht man bei der Datenerhebung von einer Stichprobe.

Urliste:
Das Ergebnis der Stichprobe wird in einer Urliste festgehalten.

Rohdaten:
Sind alle in der Urliste enthaltenen Daten.

Erhebungsumfang:
Ist die Anzahl der untersuchten Objekte (hier Schüler) Werden z.B. 27 Schüler befragt, so sagt man, “Die Anzahl der Merkmalsträger (n = 27) bildet den Erhebungsumfang“.

Merkmale:
Sind die Eigenschaften der Objekte. (z.B. Geschlecht, Körpergröße, Gewicht, Raucher, Sportart, ...)

Merkmalsausprägung x_i:
Ein Merkmal kann in verschiedenen Ausprägungen vorkommen. (z.B. Geschlecht m oder w)

Klasseneinteilung
Werden verschiedene Merkmalsausprägungen zu einer neuen Ausprägung zusammengefasst, so spricht man von einer Klasseneinteilung der Stichprobenwerte. Die Darstellung erfolgt in einem Säulendiagramm ohne Lücken.

Häufigkeiten.
Für die Merkmalsausprägung x_i gilt:

$$\text{Relative Häufigkeit von } x_i = \frac{\text{absolute Häufigkeit von } x_i}{\text{Anzahl der Merkmalsträger}} \qquad h_i = \frac{n_i}{n} \quad 0 \le n_i \le n \quad 0 \le h_i \le 1$$

Häufigkeitsdichte im Histogramm.

$$\text{Rechteckhöhe} = \frac{\text{Fläche}}{\text{Breite}} \triangleq \frac{\text{relative Klassenhäufigkeit } h_i}{\text{Klassenbreite } b_i} \quad \text{(Häufigkeitsdichte)}$$

Vergleich von Säulendiagramm und Histogramm

Säulendiagramm	Histogramm
Wenn man die relativen Häufigkeiten als Längen von Säulen veranschaulicht, entsteht ein Säulendiagramm. Die Summe der Längen aller Säulen hat den Wert 1 (100%)	Wenn man die relativen Häufigkeiten als Flächen von Rechtecken veranschaulicht, entsteht ein Histogramm. Die Summe der Flächeninhalte hat den Wert 1 (100%)

Arithmetisches Mittel einer Datenreihe

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(x_1 + x_2 + \ldots + x_n\right)$$

n: Anzahl der Beobachtungswerte x_i $\qquad$ x_i: i - ter Beobachtungswert $i \in \mathbb{N}$

Der Median (Zentralwert einer Datenreihe)

Der Median x_{Med} ist derjenige Wert (Merkmalsausprägung), der in der Mitte steht, wenn alle Beobachtungswerte x_i der Größe nach geordnet sind.

$$\frac{\text{höchstens 50\% aller B - Werte}}{\text{links vom Median}} \le \frac{\text{Median}}{\text{Median}} \le \frac{\text{höchstens 50\% aller B - Werte}}{\text{rechts vom Median}}$$

Allgemeine Rechenvorschrift zur Berechnung des Medians

Ist n ist die Anzahl der Beobachtungswerte x_i , dann gilt:

n ungerade $\Rightarrow x_{Med} = x_{\frac{n+1}{2}}$ $\qquad$ n gerade $\Rightarrow x_{Med} = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$

Modalwert.

Der Modalwert x_{Mod} ist der Merkmalswert, der am häufigsten vorkommt.

Die Spannweite.

Spannweite = größter Beobachtungswert - kleinster Beobachtungswert $\qquad R = x_{max} - x_{min}$

Quartiele und Quartilsabstand.

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
KG	1,60	1,67	**1,67**	**1,68**	1,68	1,70	**1,70**	1,72	1,73	**1,75**	**1,76**	1,78	1,84

25% $\qquad$ 25% $\qquad$ 25% $\qquad$ 25%

1. Quartil $Q_1 = 1{,}675$ $\qquad$ 2. Quartil $Q_2 = 1{,}70$ $\qquad$ 3. Quartil $Q_3 = 1{,}755$

50%

Quartilsabstand

Quartilsabstand

Der mittlere 50% - Bereich aller Beobachtungswerte heißt Quartilsabstand.

Berechnung: $\qquad Q_A = Q_3 - Q_1$

Varianz einer Datenreihe und Standardabweichung

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{\left(x_1 - \overline{x}\right)^2 + \left(x_2 - \overline{x}\right)^2 + \ldots + \left(x_n - \overline{x}\right)^2}{n}$$

n: Anzahl der Beobachtungswerte, x_i : i - ter Beobachtungswert, $\overline{x}$: Mittelwert

Standardabweichung: $s = \sqrt{s^2} = \sqrt{\text{Varianz}}$