

§ 1. Deskriptive Statistik.

Deskriptiv = beschreibend; vor allem durch Berechnung von Kennzahlen, und durch grafische Veranschaulichung

Allgemeines Setting:

Input: Daten $\mathcal{D} = \{e_1, \dots, e_N\} \subseteq \Omega$, wobei Ω eine "Menge" ist. (mehr dazu später).

Output: Informationen über \mathcal{D} .

Definition 1.1 (Grundbegriffe)

Ω heißt statistische Grundgesamtheit / Menge von Events.

\mathcal{D} heißt Stichprobe.

Beispiel: Ω = Menge aller Münzwürfe

$\mathcal{D} = \{\text{Wurf 1}, \text{Wurf 2}, \dots, \text{Wurf } N\}$

Verschiedene Merkmale von \mathcal{D} werden wie folgt modelliert:

Definition 1.2. (Merkmale)

Sei $X: \Omega \rightarrow W$ eine Funktion. (mehr zu W unten).

Wir nennen X **Merkmal / Messwert** und W **Wertebereich**

Beispiel $\Omega = \{ \text{Wurf 1, Wurf 2, ...} \}$

$\mathcal{D} = \{ \text{Wurf 1, ..., Wurf N} \}$

$W = \{ \text{Kopf, Zahl} \}$

$X =$ Funktion, die einen Wurf abbildet auf Kopf oder Zahl.

Notation $x_i := X(e_i)$ heißt **Beobachtung**

Warum der Unterschied zwischen Ω und W ? Wozu brauchen wir Ω überhaupt? Reicht es nicht nur mit Merkmalen zu arbeiten?

Antwort Wir brauchen den Unterschied zwischen Ω und W aus folgendem Grund: Wir wollen verschiedene Daten mit den gleichen Merkmalen unterscheiden! Beachte: In der Definition von $\mathcal{D}\Omega$ ist \mathcal{D} eine Menge! Und W ist eine Menge! Datenpunkte in \mathcal{D} mit den gleichen Merkmalen können in W nicht unterschieden werden. Wir brauchen \mathcal{D} um den Unterschied definieren zu können.

Wir fassen verschiedene Ausprägungen von Merkmalen zusammen.

Definition 1.3 (Merkmaltypen)

1. Wir nennen X **quantitatives Merkmal**, falls $W \subseteq \mathbb{R}$.

D.h. falls die x_i durch Angabe einer messbaren Größe definiert sind.

Wir unterscheiden dabei zwei Fälle:

diskret: der Wertebereich W ist eine diskrete Menge. D.h. W besteht aus isolierten Punkten.

Beispiel: Alter in Jahren. $\leadsto W \subseteq \mathbb{N}$.

stetig: W ist eine kontinuierliche Menge.

Beispiel: Temperatur in $^{\circ}\text{C}$ $\leadsto W = [-273, \infty)$

2. Wir nennen X **nominales Merkmal**, falls W eine endliche Menge von Wörtern oder Buchstabenfolgen ist.

(Beispiele: $W = \text{Alphabet}$. $W = \text{Wohnort}$.).

3. Wir nennen X **ordinales Merkmal**, falls es nominal ist und es zusätzlich eine Ordnung auf W gibt.

Beispiel: $W = \{ \text{"groß"}, \text{"mittel"}, \text{"klein"} \}$.

und: $\text{"klein"} < \text{"mittel"} < \text{"groß"}$.

Definition 1.4 $k \geq 2$.

Sei $X = (X_1, \dots, X_k)$ eine Liste / ein Vektor von Merkmalen,
dann nennen wir X ein **multivariates** Merkmal.

Beispiel: $X = (X_1, X_2)$, $X_1 = \text{Name einer Stadt}$
 $X_2 = \text{Anzahl Einwohner}$

diskret quantitatives nominales Merkmal