§ 1. Deskriptive Statistik.

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

Allgemeines Setting:

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

<u>Output</u>: Informationen über $\mathcal{D}$.

<u>Definition</u> 1.1 (Grundbegriffe)

Ω heißt statistische Grundgesamtheit / Menge von Events.
$\mathcal{D}$ heißt Stichprobe.

<u>Beispiel</u>: Ω = Menge aller Münzwürfe
$\mathcal{D} = \{$Wurf 1, Wurf 2, ..., Wurf N$\}$

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

Definition 1.2. (Merkmale)

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

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

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

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

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

$X =$ 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!** Bemerkung: In der Definition von $D \subset \Omega$ ist D eine Menge! Und W ist eine Menge! Datenpunkte in D mit den gleichen Merkmalen können in W nicht unterschieden werden. Wir brauchen D um den Unterschied definieren zu können.

Wir fassen verschiedene Ausprägungen von Merkmalen zusammen.

<u>Definition 1.3</u> (Merkmalstypen)

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:
   <u>diskret</u>: der Wertebereich W ist eine diskrete Menge. D.h. W besteht aus isolierten Punkten.
   <u>Beispiel</u>: Alter in Jahren. $\rightsquigarrow$ $W \subseteq \mathbb{N}$.
   <u>stetig</u>: W ist eine kontinuierliche Menge.
   <u>Beispiel</u>: Temperatur in °C $\rightsquigarrow$ $W = [-273, \infty)$

2. Wir nennen X nominales Merkmal, falls W eine endliche Menge von Wörtern oder Buchstabenfolgen ist.
   (<u>Beispiel</u>: W = Alphabet. W = Wohnort.).

3. Wir nennen X ordinales Merkmal, falls es nominal ist <u>und</u> es zusätzlich eine Ordnung auf W gibt.
   <u>Beispiel</u> $W = \{$"groß", "mittel", "klein"$\}$.
   und: "klein" < "mittel" < "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 = Name einer Stadt

X_2 = Anzahl Einwohner

diskret quantitatives nominales Merkmal