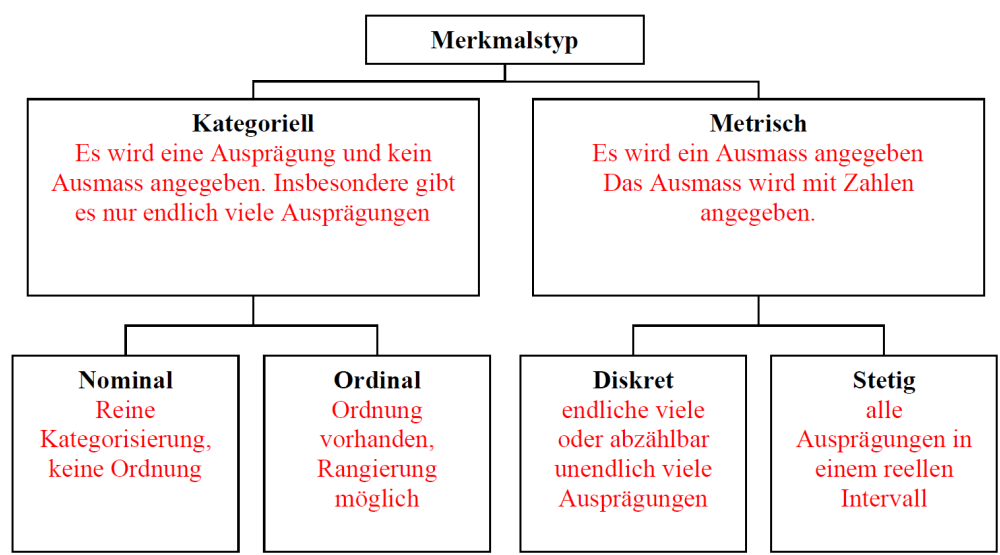


# Stichproben

Eine *Stichprobe*  $(x_1, \dots, x_n)$  der Länge  $n$  ist eine Auswahl von  $n$  Objekten aus einer Grundgesamtheit. Meist handelt es sich um eine *Zufallsstichprobe*, d.h. die Auswahl der Objekte ist zufällig und repräsentativ. Relevant ist oft nur ein bestimmtes *Merkmal*, z.B. die Partei, für welche die i-te befragte Person in der kommenden Wahl stimmen wird. Das untersuchte Merkmal kann bestimmte *Ausprägungen* annehmen. Im Beispiel wäre das die Menge der zur Wahl stehenden Parteien. Eine solche Stichprobe  $(x_1, \dots, x_n)$  wäre zum Beispiel eine Folge von  $n$  Parteinamen  $x_i$ .

Merkmale, wie die Partei, die Augenfarbe, das Alter, die Prüfungsnote, die Laufzeit, etc. werden danach eingeteilt, ob die möglichen Ausprägungen eine natürliche Ordnungstruktur tragen oder ob nur ein nominaler Zusammenhang besteht. Im obigen Beispiel handelt es sich um eine Menge von Parteinamen, welche ohne weiteres zutun keine Anordnung besitzen. Es handelt sich hier um einen nominalen Merkmalstyp. Allgemein werden die folgenden 4 *Merkmalstypen unterschieden*:



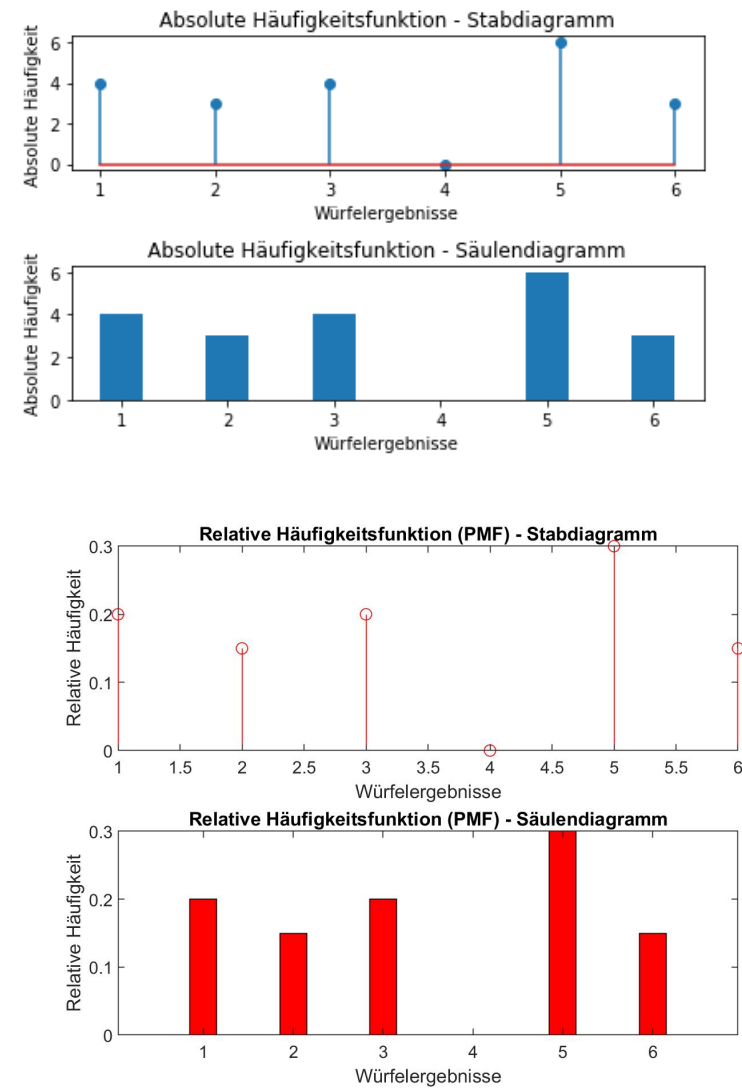
## Absolute und relative Häufigkeitsfunktionen

Sei  $(x_1, \dots, x_n)$  eine Stichprobe eines metrischen Merkmals.

Die reelle Funktion  $h : \mathbb{R} \rightarrow \mathbb{R}$ ,  $h(x) = \text{Anzahl der } 1 \leq i \leq n \text{ mit } x_i = x$ , heisst (*empirische*) *absolute Häufigkeitsfunktion* der Stichprobe.

Die reelle Funktion  $f : \mathbb{R} \rightarrow [0, 1]$ ,  $f(x) = \frac{h(x)}{n}$ , heisst (*empirische*) *relative Häufigkeitsfunktion* oder (*empirische*) *Dichtefunktion* der Stichprobe. Häufig wird dafür das Kürzel *PMF* vom Englischen probability mass function, benutzt.

Visualisiert werden die beiden Funktionen als Stab-oder Säulendiagramm. Nachfolgend Stab-und Säulendiagramme für ein Würfelexperiment zu sehen.



## Absolute und relative Summenhäufigkeit für nichtklassierte Stichproben

Sei  $(x_1, \dots, x_n)$  eine Stichprobe eines metrischen Merkmals mit der absoluten Häufigkeitsfunktion  $h(x)$  und der PMF  $f(x)$ .

Die reelle Funktion  $H : \mathbb{R} \rightarrow \mathbb{R}$ ,  $H(x) = \text{Summe aller } h(y) \text{ mit } y \leq x$  heisst *(empirische) absolute Summenhäufigkeit* der Stichprobe.

Die reelle Funktion  $F : \mathbb{R} \rightarrow [0, 1]$ ,  $F(x) = \text{Summe aller PMF Werte } f(y) \text{ mit } y \leq x$  heisst *(empirische) relative Summenhäufigkeit oder (empirische) kumulative Verteilungsfunktion* oder kurz *CMF*, vom Englischen cumulative mass function, der Stichprobe.

In den Diagrammen sind die absolute Summenhäufigkeit und die CMF aus einem Beispiel zu sehen.

