

# Statistik

CH.2 - Begriff der Häufigkeit

SS 2022 | | Prof. Dr. Buchwitz, Sommer, Henke

Wir geben Impulse

- Ziel 1
- Ziel 2
- Ziel 3

## Definition: Absolute Häufigkeit

Beobachtet man an  $n$  statistischen Einheiten ein Merkmal  $X$  mit  $k$  Merkmalsausprägungen, so gilt für die **absolute Häufigkeit**  $H_j$  der  $j$ -ten Merkmalsausprägung.

$$H_j = \text{"Anzahl der Beobachtungswerte, die in der } j\text{-ten Ausprägung auftreten" für } j = 1, \dots, k$$

## Definition: Relative Häufigkeit

Die **relative Häufigkeit**  $h_j$  gibt den prozentualen Anteil der statistischen Einheiten, die die  $j$ -te Merkmalsausprägung tragen, an.

$$h_j = \frac{1}{n} \cdot H_j \quad \text{für } j = 1, \dots, k$$

# Beispiel: Absolute und relative Häufigkeit

```
# Daten: Alter befragter Personen  
age <- c(19, 21, 20, 21, 19, 23, 22, 19, 20, 19)
```

```
# Absolute Häufigkeit  
table(age)
```

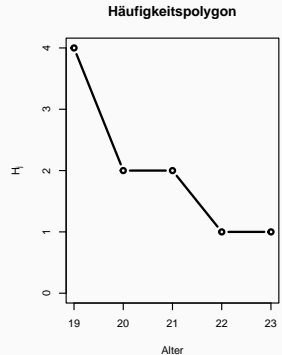
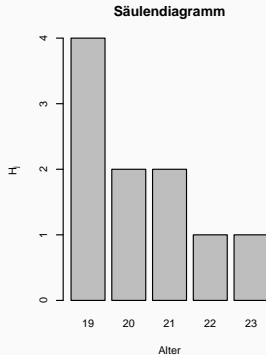
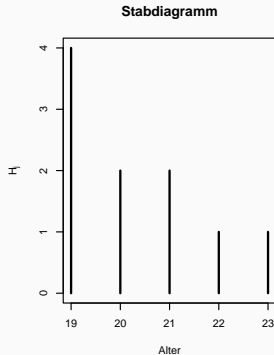
```
## age  
## 19 20 21 22 23  
##  4  2  2  1  1
```

```
# Relative Häufigkeit  
proportions(table(age))
```

```
## age  
##  19  20  21  22  23  
## 0.4 0.2 0.2 0.1 0.1
```

# Beispiel: Graphische Darstellungsmöglichkeiten von Häufigkeiten

```
par(mfrow=c(1,3))
plot(table(age), main = "Stabdiagramm",
     xlab = "Alter", ylab = expression(H[j]))
barplot(table(age), main = "Säulendiagramm",
         xlab = "Alter", ylab = expression(H[j]))
plot(table(age), type="b", main = "Häufigkeitspolygon",
     xlab = "Alter", ylab = expression(H[j]))
```



- Ein Histogramm ist eine Möglichkeit, die Häufigkeitsverteilung klassierter Daten graphisch darzustellen.
- Die Häufigkeiten werden als aneinander stoßende Rechtecke dargestellt, deren Flächen proportional zur Häufigkeit der Beobachtungswerte der Klassen sind.
- Nicht die Höhe des Rechteckes über eine Klasse ist proportional zur Klassenhäufigkeit, sondern die Fläche selbst.
- Erstellt werden können Histogramme mit dem R-Befehl `hist()`.

- In einer Befragung wurden  $n = 80$  Personen nach der Anzahl der verbrachten Urlaubstage pro Jahr befragt. Die Daten wurden in Klassen unterschiedlicher Breiten erfasst und die jeweiligen Häufigkeiten ausgezählt bzw. berechnet.

| $j$ | Klasse $K_j$ | $H_j$ | $h_j$ | Klassenbreite | norm. $H_j$ | norm. $h_j$ |
|-----|--------------|-------|-------|---------------|-------------|-------------|
| 1   | [0;8)        | 6     | 0.075 | 8             | 0.750       | 0.0094      |
| 2   | [8;18)       | 16    | 0.200 | 10            | 1.600       | 0.0200      |
| 3   | [18;25)      | 20    | 0.250 | 7             | 2.857       | 0.0357      |
| 4   | [25;30)      | 14    | 0.175 | 5             | 2.800       | 0.0350      |
| 5   | [30;35)      | 12    | 0.150 | 5             | 2.400       | 0.0300      |
| 6   | [35;43)      | 12    | 0.150 | 8             | 1.500       | 0.0187      |

Die normierten absoluten und relativen Klassenhäufigkeiten setzen die Klassenhäufigkeiten in Relation zur Klassenbreite, sodass diese Werte einen Vergleich erlauben. Es gilt  $\text{norm. } H_j = H_j / \text{Klassenbreite}$  (norm.  $h_j$  analog).

# Beispiel: Histogramm

Histogramm der klassifizierten Urlaubstage



Die Fläche der zu den Klassen 5 und 6 gehörenden Rechtecke ist identisch, da gilt  $H_5 = H_6$  (bzw.  $h_5 = h_6$ ).



## Definition: Empirische Verteilungsfunktion

Die empirische Verteilungsfunktion  $S$  eines an  $n$  Untersuchungseinheiten beobachteten Merkmals  $X$ , das  $k$  Merkmalsausprägungen  $a_1, \dots, a_k$  besitzt, wird durch die Folge der relativen Summenhäufigkeiten bestimmt.

$$S_j = \frac{1}{n} \sum_{i=1}^j H_j = \sum_{i=1}^j h_j \quad \text{für } j = 1, \dots, k$$

- Fortsetzung auf Basis der Befragung zum Urlaubsverhalten von  $n = 80$  Personen.
- TODO: Tabelle

| $j$ | Klasse $K_j$ | $H_j$ | $h_j$ | emp. Verteilungsfkt. |
|-----|--------------|-------|-------|----------------------|
| 1   | $[0;8)$      | 6     | 0.075 | 0.075                |
| 2   | $[8;18)$     | 16    | 0.200 | 0.275                |
| 3   | $[18;25)$    | 20    | 0.250 | 0.525                |
| 4   | $[25;30)$    | 14    | 0.175 | 0.700                |
| 5   | $[30;35)$    | 12    | 0.150 | 0.850                |
| 6   | $[35;43)$    | 12    | 0.150 | 1.000                |

Empirische Verteilungsfunktion



- Wodurch unterscheiden sich absolute von relativen Häufigkeiten?
- Was ist der Unterschied zwischen einem Stabdiagramm und einem Histogramm?
- Was zeigt die empirische Verteilungsfunktion?

- Wodurch unterscheiden sich absolute von relativen Häufigkeiten?
  - ▶ Die absolute Häufigkeit gibt die Anzahl der Beobachtungen wieder, die relative Häufigkeit gibt den Anteil der Beobachtung an der Gesamtanzahl wieder.
- Was ist der Unterschied zwischen einem Stabdiagramm und einem Histogramm?
  - ▶ In einem Stabdiagramm gibt die Höhe des Stabs die Häufigkeit wieder, in einem Histogramm ist es die Fläche der Balken, die die Häufigkeit wiedergibt.
- Was zeigt die empirische Verteilungsfunktion?
  - ▶ Anhand der empirischen Verteilungsfunktion kann man die relative Häufigkeit ablesen, mit der ein Wert höchstens eine bestimmte Größe hat. Die empirische Verteilungsfunktion gibt so die kumulierten relativen Häufigkeiten bis zu einem bestimmten Wert an.