

Unterrichtsmaterialien zur Lehrveranstaltung Mathematik für Biologen

Häufigkeiten:

Gegeben sei eine Grundgesamtheit (häufig Population genannt) von n Objekten. Wir interessieren uns dafür, wie oft die Ausprägungen a_1, a_2, \dots, a_k eines bestimmten Merkmals in der Grundgesamtheit vorkommen.

Nach Auszählen (z.B. durch Strichlisten) erhalten wir die (absoluten) Häufigkeiten h_1, h_2, \dots, h_k ("5 Bäume sind Eichen, 3 Bäume sind Fichten, 2 Bäume sind Buchen") des Auftretens der Ausprägungen in der Grundgesamtheit, woraus sich die relativen Häufigkeiten $r_1 = \frac{h_1}{n}, r_2 = \frac{h_2}{n}, \dots, r_k = \frac{h_k}{n}$ ergeben ("5/10 der Bäume sind Eichen, 3/10 der Bäume sind Fichten, 2/10 der Bäume sind Buchen").

Es ist $0 \leq h_i \leq n, 0 \leq r_i \leq 1$ für $i = 1, 2, \dots, k$ sowie

$$h_1 + h_2 + \dots + h_k = n,$$

$$r_1 + r_2 + \dots + r_k = 1.$$

(Kurzschreibweisen: $\sum_{i=1}^k h_i = n, \sum_{i=1}^k r_i = 1$)

Ist $A \subset M := \{a_1, a_2, \dots, a_k\}$ eine Teilklasse der Ausprägungen eines Merkmals ("Laubbäume"), so bezeichnet man mit $h(A)$ die absolute Häufigkeit, mit der mindestens eine Ausprägung aus A in der Grundgesamtheit auftritt ("8 Bäume sind entweder Eichen oder Buchen"). $r(A) := \frac{h(A)}{n}$ ist die relative Häufigkeit, mit der in der Grundgesamtheit mindestens eine Ausprägung aus A auftritt ("8/10 der Bäume sind Laubbäume").

Rechenregeln

Für disjunkte Mengen $A, B \subset M := \{a_1, a_2, \dots, a_k\}$ gilt:

$$h(A) + h(B) = h(A \cup B), \quad h(M) = n$$

$$r(A) + r(B) = r(A \cup B), \quad r(M) = 1$$

Speziell: $r(A) + r(A^c) = 1$.

Graphische Darstellung von Häufigkeiten

a) Stabdiagramm (Balkendiagramm)

Die graphische Darstellung von absoluten oder relativen Häufigkeiten durch Stäbe (durch Balken) über den Ausprägungen nennt man Stabdiagramm (oder auch Balkendiagramm).

b) Histogramm

Gegeben seien n reelle Zahlen x_1, x_2, \dots, x_n als Realisierungen eines kontinuierlichen Merkmals (z.B. Flügellänge).

i) Einfache Version

Ausgehend vom zugehörigen Stabdiagramm werden die Stäbe der Häufigkeiten der möglichen Ausprägungen dieses Merkmals verbreitert, bis sie aneinanderstoßen.

Wichtig: die Fläche über jeder Ausprägung muß proportional zur Häufigkeit dieser Ausprägung sein.

ii) Übliches Histogramm:

Man zerlege den Bereich zwischen kleinster und größter Reasisierung in l gleichgroße Teilintervalle I_1, I_2, \dots, I_l (üblich: l ungefähr gleich \sqrt{n}).

Dann ermittelt man die Häufigkeiten h'_1, h'_2, \dots, h'_l , mit denen die n Zahlen x_1, x_2, \dots, x_n in diese Intervalle fallen und stellt diese Häufigkeiten graphisch dar wie unter i) beschrieben.