Verteilungen und ihre Darstellung

Häufigkeitsverteilungen

Sei $\Omega = \{x_1, x_2, \ldots, x_m\}$ der Merkmalraum irgend einer Zufallsvariablen. Liefert sie in einer Versuchsreihe die Stichprobe S, so nennen wir die Folge der relativen Häufigkeiten $h_i = H_S(x_i)$ $(i = 1, 2, \ldots, m)$ **Häufigkeitsverteilung von** S (über dem Merkmalraum Ω); wir notieren sie auch als m-Tupel: $(h_1, h_2, \ldots, h_m)$.

Der folgende Satz kennzeichnet die m-Tupel (endlichen Folgen) reeller Zahlen, die Häufigkeitsverteilungen (H-Verteilungen) einer Stichprobe sind:

Satz

$(h_1, h_2, \ldots, h_m) \in \mathbb{R}^m$ ist genau dann eine Häufigkeitsverteilung einer Stichprobe S, wenn gilt:

(1) $h_i \geq 0$ für alle $i \in \{1, 2, \ldots, m\}$

(2) $h_1 + h_2 + \ldots + h_m = 1$

(3) Es gibt eine ganze Zahl s derart, dass alle Produkte $s \cdot h_i$ ganz sind ($1 \leq i \leq m$).

Beweis:

1. Es ist unmittelbar einleuchtend, dass eine H-Verteilung die Eigenschaften (1) und (2) erfüllt. Da die h_i rationale Zahlen sind, kann s als das kleinste gemeinschaftliche Vielfache der Nenner in einer Darstellung sämtlicher h_i als Brüche gewählt werden, womit auch (3) gezeigt ist.

2. Seien nun umgekehrt die Eigenschaften (1)-(3) vorausgesetzt. Wir setzen $\Omega = \{1, 2, \ldots, m\}$ und denken uns eine Stichprobe S gegeben, in welcher das Merkmal i genau $s \cdot h_i$-mal vorkommt (leicht zu erreichen). Dann gilt: $|S| = \sum_{i=1}^{m} s \cdot h_i = s(h_1 + \ldots + h_m) = s$ sowie $H_S(i) = \frac{s \cdot h_i}{s} = h_i \geq 0$ für alle $i \in \{1, 2, \ldots, m\}$. Das vorgegebene m-Tupel $(h_1, h_2, \ldots, h_m)$ ist somit die H-Verteilung der so konstruierten Stichprobe S. ■

■ Verallgemeinerung

Der Begriff der H-Verteilung lässt sich noch etwas allgemeiner fassen, wenn die im obigen Satz genannte Bedingung (3) entfällt sowie zusätzlich erlaubt wird, dass die Zahlenfolge unendlich ist.

Eine Folge reeller Zahlen $(p_1, p_2, \ldots p_n, \ldots)$ heiße (diskrete) **Verteilung**, wenn gilt:

1. $p_i \geq 0$ für alle $i \geq 1$
2. $\sum_{i \geq 1} p_i = 1$.

Verteilung wovon?

Lassen sich die Zahlen p_i als relative Häufigkeiten elementarer Merkmale deuten, so haben wir es mit einer H-Verteilung zu tun.

Lassen die p_i sich als Wahrscheinlichkeiten von Elementarereignissen deuten, so sprechen wir von einer Wahrscheinlichkeitsverteilung (W-Verteilung).

Z.B. benötigt man bei Entscheidungen unter Risiko eine W-Verteilung über der Menge B der Bedingungen.

Darstellungsformen

Für Stichproben beobachteter Merkmale und für ihre H-Verteilungen gibt es eine Reihe unterschiedlicher Darstellungsformen. Welcher Typ von Diagramm in einer bestimmten Situation angemessen ist, hängt vom Zweck der Visualisierung und von der jeweils vorliegenden Ausprägung der Merkmale ab.

Wir behandeln hier nur kurz die wichtigsten Diagrammformen für Stichproben und für H-Verteilungen.

Darstellung von Stichproben:

Punktediagramme

Liniendiagramme

Darstellung von Verteilungen:

Stabdiagramme

Kreisdiagramme

Häufigkeitspolygone

■ Computergestützte Visualisierung

Die Darstellungsformen werden im Folgenden mit Hilfe eines *Mathematica*-Pakets "Verteilungen.m" demonstriert, das es u.a. erlaubt, Zufallsstichproben über beliebigen endlichen Merkmalräumen zu erzeugen und die H-Verteilungen numerisch und grafisch auszugeben.

Zunächst muss das Paket geladen werden:

```
<< Graphics`Graphics`
<< Modellbildung`Verteilungen`
```

Wir legen als Merkmalraum die Menge der Augenzahlen eines Spielwürfels zu Grunde und erzeugen über ihm eine Zufallsstichprobe vom Umfang 30:

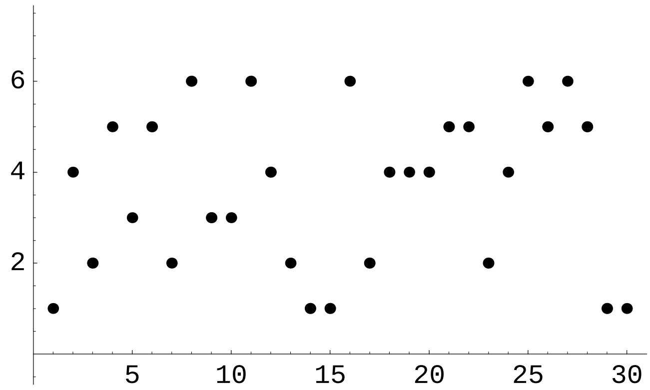
```
m = {1, 2, 3, 4, 5, 6};
sprobe = Stichprobe[m, 30]
```

```
{1, 4, 2, 5, 3, 5, 2, 6, 3, 3, 6, 4, 2,
 1, 1, 6, 2, 4, 4, 4, 5, 5, 2, 4, 6, 5, 6, 5, 1, 1}
```

■ Punktediagramme

In einem rechtwinkligen Koordinatensystem gezeichnete Punktmengen. Auf der Ordinate werden die Merkmale abgetragen. Ausprägungen: quantitativ, ordinal (unter Umständen auch nominale Merkmale möglich). Die Abszisse dient i.a. als Zeitachse, d.h. die Merkmale treten in ihrer zeitlichen Reihenfolge auf:

```
PunkteDiagramm[sprobe, 2]
```

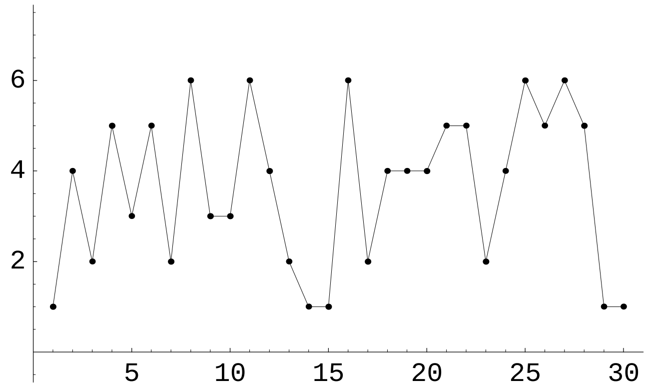


Hier ist die Stichprobe mit der Punktgröße 2 dargestellt (mögliche Punktgrößen: 1, 2, 3).

■ Liniendiagramme

In einem rechtwinkligen Koordinatensystem gezeichnete Kurven (Linien). Die Abszisse ist Zeitachse, auf der Ordinate werden die Merkmale abgetragen (nur quantitative Merkmale sinnvoll!). Im Prinzip wird der Merkmalraum als kontinuierlich angenommen. Die Verbindung der Datenpunkte durch Kurvenstücke (häufig: Strecken) ermöglicht es, Zwischenwerte zu schätzen bzw. Trends zu erkennen. Liniediagramme betonen den Gesamtverlauf bzw. den Zusammenhang, der (eventuell) zwischen den Beobachtungsdaten besteht.

```
LinienDiagramm[sprobe, 1]
```



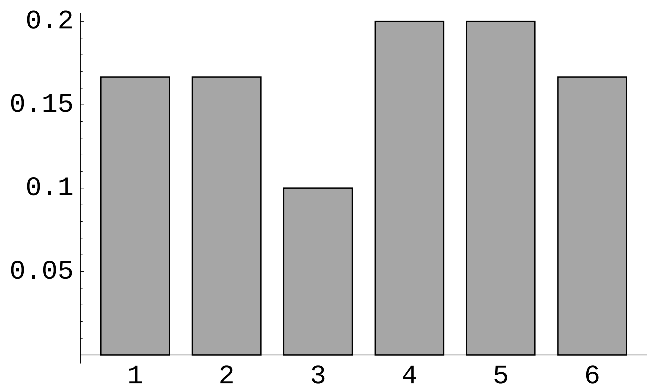
Hier ist die Stichprobe mit der Liniendicke 1 dargestellt (mögliche Liniedicken: 1, 2, 3).

■ Stabdiagramme

Den Begriff "Stabdiagramm" verwenden wir hier synonym zu "Histogramm". Stabdiagramme visualisieren eine Verteilung aus relativen Häufigkeiten. (Eine Vorform des Stabdiagramms, das sog. Säulendiagramm, besteht aus rechteckigen Säulen, deren Höhen die absoluten Häufigkeiten von Merkmalen in einer Stichprobe wiedergeben.)

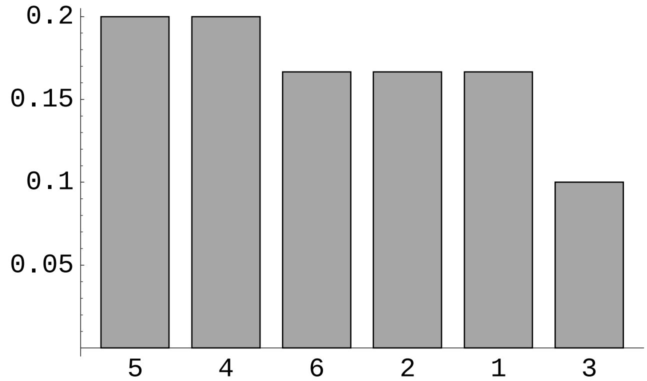
Es liegt ein rechtwinkliges Koordinatensystem zugrunde. Auf der Abszisse werden die Elemente des Merkmalraums abgetragen (alle Ausprägungen möglich!) und über ihnen je ein schmales Rechteck gezeichnet, dessen Höhe die relative Häufigkeit des Merkmals darstellt.

```
StabDiagramm[m, sprobe]
```



Nützlich ist (besonders im Fall nominaler Merkmale) das **Staffeldiagramm** (angeordnetes Stabdiagramm), das die Merkmale nach abnehmender (bzw. zunehmender) Häufigkeit darstellt:

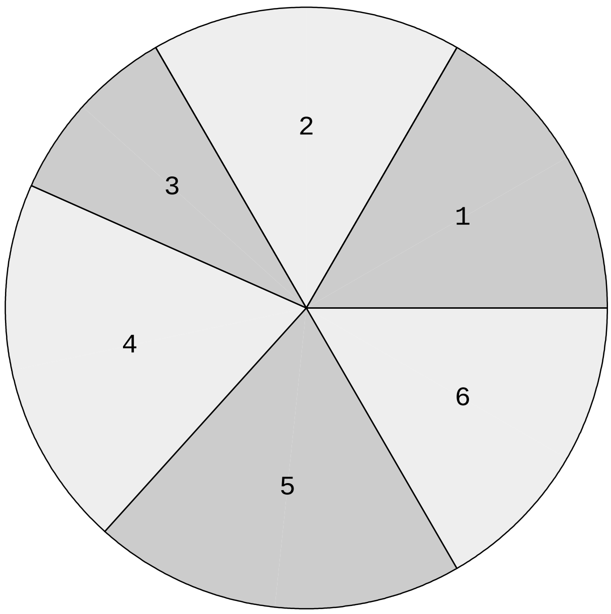
```
StaffelDiagramm[m, sprobe]
```



■ Kreisdiagramme

Kreisdiagramme (auch "Tortendiagramme" genannt) stellen eine Verteilung dar, indem sie jedem Merkmal einen Kreissektor ("Kuchenstück") zuordnen, dessen Anteil am Kreis der relativen Häufigkeit des betreffenden Merkmals entspricht. Diese Darstellungsart eignet sich besonders für nominale Merkmale. Grundsätzlich bietet sie sich aber immer dann an, wenn die Aufteilung eines Ganzen in Anteile gezeigt bzw. betont werden soll.

```
KreisDiagramm[m, sprobe]
```



■ Häufigkeitspolygone

Es handelt sich um Streckenzüge in einem rechtwinkligen Koordinatensystem. Die Strecken verbinden Punkte, deren Abszissenwerte quantitative Merkmale (häufig Messwerte) und deren Ordinatenwerte die zugehörigen relativen Häufigkeiten (in der Stichprobe bzw. Messreihe) sind.

Der Merkmalraum ist bevorzugt als kontinuierlich vorzustellen, sodass die Punkte auf den Strecken als Zwischenwerte gedeutet werden können.

```
HaeufigkeitsPolygon[m, sprobe]
```