Statistik

Die Statistik gliedert sich in

- die **Beschreibende Statistik**, die Daten erfasst und diese durch Tabellen, Graphiken und Kennzahlen, möglichst übersichtlich beschreibt;
- die **Beurteilende Statistik**, die auf Basis der Beschreibenden Statistik prognostiziert und vergleicht, sich also z. B. mit der Qualitätskontrolle von Produkten beschäftigt.

Arbeitsweise der Statistik, Datenerhebung

Wenn die Datenerhebung speziell für die statistische Untersuchung erfolgt spricht man von einer sogenannten **Primärerhebung**, zum Unterschied von der **Sekundärerhebung**, bei der das Datenmaterial schon vorliegt.

Die **Grundgesamtheit** ist die Menge der zu beurteilenden Objekte; der **Umfang der Grundgesamtheit** ist die Anzahl ihrer Elemente. Eine **Stichprobe** ist eine Menge von Objekten, die einer Grundgesamtheit zufällig (d.h. mit gleicher Chance für jedes Objekt der Grundgesamtheit) entnommen werden. Der **Stichprobenumfang** ist die Anzahl der Elemente einer Stichprobe.

Die Arbeitsweise der Statistik scheint simpel zu sein: Zunächst werden die Daten erfasst, dann erfolgt die Datenaufbereitung, d.h. das Ordnen und Verdichten der Daten in Form von Tabellen bzw. Diagrammen. Schließlich wird noch die Schlussfolgerung gezogen — und die Sache hat sich. Leider ist es in der Praxis keineswegs ganz so einfach.

Merkmal, Merkmalsträger, Merkmalsausprägung

Ein **Merkmal** ist eine Eigenschaft, die zur Beurteilung der zu untersuchenden Objekte **(Merkmalsträger)** dienen kann. Unter einer **Merkmalsausprägung** versteht man einen Wert (eine Eigenschaft), den ein Merkmal annimmt.

Merkmalsausprägungen sind **quantitativer Art**, wenn sie nur durch Zahlen dargestellt werden können, andernfalls sind sie **qualitativer Art**. Bei quantitativen Merkmalen unterscheidet man außerdem zwischen **stetigen** und **diskreten** Merkmalsausprägungen. Stetige Ausprägungen können jeden beliebigen Wert eines bestimmten Intervalls der Menge $\mathbb{R}$ annehmen. Diskrete Ausprägungen können im betrachteten Intervalls nur endlich viele Zahlenwerte annehmen, daher sind nur spezielle Werte möglich die z.B. durch Zählen ermittelt werden.

Beispiel: Die Laufgeschwindigkeit einer bestimmten Klasse variiert zwischen 10 und 18 km/h. Das ist ein Beispiel für eine quantitative, stetige Merkmalsausprägung.

Familienbesitzer in Österreich besitzen zwischen 1. und 15. Kinder. Das ist ein Beispiel für für eine quantitative, diskrete Merkmalsausprägung (durch Zählen ermittelbar).

Absolute und relative Häufigkeit

Die Anzahl der Stichprobenwerte von ein und derselben Merkmalsausprägung heißt **absolute Häufigkeit** dieser Merkmalsausprägung und wird mit H_i bezeichnet.

Das Verhältnis der absoluten Häufigkeit H_i einer Merkmalsausprägung zum Gesamtumfang n der Stichprobe heißt **relative Häufigkeit** und wird mit h_i bezeichnet.

$$\text{Relative Häufigkeit} = \frac{\text{absolute Häufigkeit}}{\text{Stichprobenumfang}}$$

$$h_i = \frac{H_i}{n}$$

Klasseneinteilung

Unter **Klasseneinteilung** versteht man die Zusammenfassung zweier oder mehrer benachbarter Merkmalsausprägungen zu Gruppen oder Klassen.

Die Klasseneinteilung hat Vor- und Nachteile:

- Je weniger Klassen desto übersichtlicher; die Information wird aber gleichzeitig kleiner.
- Je mehr Klassen, desto unübersichtlicher; es treten aber Eigentümlichkeiten der Häufigkeitsverteilung hervor, die bei weniger Klassen nicht zu erkennen wären.

Was versteht man unter Mittelwert?

Das arithmetische Mittel ist der bekannteste Mittelwert:

Das **arithmetische Mittel** $\overline{x}$ (gesprochen : x quer) der Zahlen $x_1, x_2, \ldots, x_n$ erhält man, indem man die Summe dieser Zahlen durch ihre Anzahl n dividiert.

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Das arithmetische Mittel ist ein wichtiger, aber nicht der einzige Mittelwert; einige Beispiele für andere Mittelwerte:

- Das gewogene arithmetische Mittel
- Das geometrische Mittel
- Der Zentralwert
- Der Modalwert

Gewogenes arithmetische Mittel $\overline{x}$:

Beim gewogenen arithmetischen Mittel wird der Mittelwert mit dem Umfang der betreffenden Gruppe multipliziert:

$$\overline{x} = \frac{\overline{x}_1 m_1 + \overline{x}_2 m_2 + ... + \overline{x}_n m_n}{m_1 + m_2 + ... + m_n} = \frac{\sum\limits_{i=1}^{n} \overline{x}_i m_i}{\sum\limits_{i=1}^{n} m_i}$$

$\overline{x}_1$, $\overline{x}_2$ ... Mittelwerte
m_1, m_2 ... Umfänge der betreffenden Gruppen

Geometrisches Mittel $\overline{x}_g$:

Das geometrische Mittel wird dann verwendet, wenn der Durchschnitt von positiven oder negativen Wachstumsraten berechnet werden soll.

$$\overline{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Zentralwert z:

Ordnet man eine Liste x_1, x_2, ..., x_n der Größe nach so heißt derjenige Wert, der die mittlere Lage aller Werte einnimmt, Zentralwert z oder Median. Der Zentralwert halbiert also die gesamte Liste.
Bei einer Geraden Anzahl von Werten wird der Mittelwert der beiden in der Mitte stehenden Zahlen genommen.

Modalwert m:

In einer Liste von Merkmalsausprägungen heißt der Wert, der bezüglich der anderen Ausprägungen am häufigsten vorkommt, **Modalwert** m.

Streuungsmaße

Der Mittelwert allein sagt wenig über eine Verteilung aus. Mit Hilfe der sogenannten Streuung versucht man die Abweichungen der einzelnen Datenwerte vom Mittelwert zu beschreiben.

Spannweite w: Die Differenz zwischen der kleinsten und der größten Merkmalsausprägung einer Verteilung bezeichnen wir als **Spannweite** w.

Mittlere lineare Abweichung e: Das arithmetische Mittel aus den Absolutbeträgen der Abweichungen aller Merkmalsausprägungen von deren arithmetischem Mittel bezeichnen wir als **mittlere lineare Abweichung** e.

Die Definition ist nicht ganz leicht zu verstehen, hier ein Beispiel zum Verständnis:
Die Daten 5, 9, 1 sind gegeben – das arithmetische Mittel dieser Zahlen ist 5 (leicht nachzurechnen). Berechnen wir nun die Absolutbeträge der Abweichung der Daten vom arithmetischen Mittel (in dem Fall von 5) : $|5 - 5| = 0$; $|9 - 5| = 4$; $|1 - 5| = 4$.

Das arithmetische Mittel der Zahlen 0, 4, 4 ist schließlich 2.6. 2.6 ist die mittlere lineare Abweichung der Zahlen 5, 9 und 1!

$$e = \frac{|x_1 - \overline{x}| + \cdots + |x_n - \overline{x}|}{n} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}|$$

Bei der mittleren linearen Abweichung e wird den Abweichungen aller Werte vom Mittelwert die gleiche Bedeutung beigemessen, unabhängig vom Ausmaß der Abweichung. Durch Quadrieren der Differenzen kann aber erreicht werden, dass das Streuungsmaß größere Abweichungen auch entsprechend stärker berücksichtigt als kleinere. Das so errechnete Streuungsmaß wird als **Varianz** s^2 bezeichnet.

$$s^2 = \frac{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n} = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$$

Standardabweichung s: Die Standardabweichung kann man als das wichtigste Streuungsmaß betrachten.

$$s = \sqrt{\frac{\text{Summe der quadratischen Abweichungen vom Mittelwert}}{\text{Anzahl der Stichprobenwerte}}} = \sqrt{\frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n}}$$