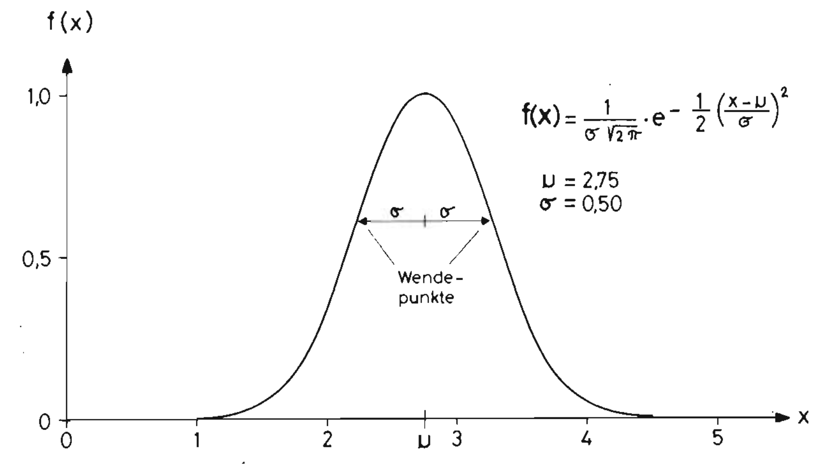
Normalverteilung

Kurve zur
Darstellung der Häufigkeit
in einer Grundgesamtheit

In der Statistik versteht man unter einer theoretischen Verteilung die formelmäßige Angabe oder die graphische Darstellung der Wahrscheinlichkeit, anschaulich realisiert durch die relative Häufigkeit, einer vom Zufall abhängigen Variablen in Abhängigkeit von den Werten, die die Zufallsvariable annehmen kann. Eine theoretische Verteilung wird mathematisch abgeleitet und ist zunächst einmal ein abstrakter Formalismus. Das reale Gegenstück zu einer theoretischen Verteilung ist die empirische Verteilung, zum Beispiel die graphische Darstellung der Häufigkeit der in einer Grundgesamtheit oder einer Stichprobe auftretenden Werte, etwa die Häufigkeitsverteilung der Körpergewichte von Normalpersonen.



Darstellung 1: Die Normalverteilung wird graphisch durch die „Glockenkurve" dargestellt. Sie gibt für jeden Wert der Variablen x die Häufigkeit an, mit der dieser Wert in der Grundgesamtheit vorkommt

Bedeutung

Die wichtigste theoretische Verteilung ist die von Gauß bei der Untersuchung der Meßfehler eingeführte Normalverteilung oder Gaußverteilung, weil

❶ viele in der Praxis vorkommende Zufallsvariablen normalverteilt sind,

❷ viele weitere Zufallsvariablen annähernd normalverteilt sind und man unter Zugrundelegung einer Normalverteilung bei statistischen Rechnungen und Schlüssen für die Praxis brauchbare Ergebnisse erhält,

❸ manche nicht normalverteilte Variablen durch eine einfache Umrechnung, zum Beispiel eine Logarithmierung, in eine Normalverteilung überführt werden können und

❹ kompliziertere Verteilungen in Grenzfällen gut durch eine Normalverteilung angenähert werden können.

Darstellung

Die formelmäßig durch den relativ kompliziert erscheinenden Ausdruck

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

– mit den Konstanten $\pi = 3{,}142$ und $e = 2{,}178$ – beschriebene Normalverteilung wird graphisch durch die allgemein bekannte Glockenkurve dargestellt (Darstellung 1). Diese Kurve gibt für jeden Wert der Variablen x die Häufigkeit f (x), mit der dieser Wert in einer großen – theoretisch unendlich großen – Gesamtheit vorkommt. (Exakt muß, weil es sich um eine stetige Verteilung handelt, anstelle eines Wertes x ein kleines Intervall Δx betrachtet werden).

Gekennzeichnet wird eine Normalverteilung durch:

❶ ihre Lage auf der x-Achse, die durch den Parameter μ in der Formel angegeben wird. Bei $x = \mu$ liegt das Maximum der Kurve; dieser Wert kommt am häufigsten in der statistischen Gesamtheit vor; es ist der Mittelwert der Verteilung;

❷ ihre Breite in der Höhe der Wendepunkte, die in die Formel über σ eingeht. Es ist die Standardabweichung der Verteilung. Je kleiner σ ist, desto schmäler ist die Kurve, desto steiler fällt sie ab, aber auch desto höher ist das Maximum, desto weniger streuen die Einzelwerte um den Mittelwert μ.

Durch die Gestalt der Formel wird die Kurve so normiert, daß der Flächeninhalt zwischen der x-Achse und der Kurve immer gleich ist. Deshalb ist die Normalverteilung durch die Parameter μ und σ hinsichtlich ihrer Lage auf der x-Achse und ihrer Form auch vollständig bestimmt.

Durch eine einfache Umrechnung kann man dafür sorgen, daß jede in der Praxis auftretende Normalverteilung, für die ja im allgemeinen zunächst $\mu \neq 0$ und $\sigma \neq 1$ ist, in die sogenannte Standardnormalverteilung mit dem Mittelwert $\mu = 0$ und der Standardabweichung $\sigma = 1$ transformiert wird.

Anwendungen

Die Werte dieser Standardnormalverteilung N (0; 1) sind in Tafeln angegeben, die jedes Statistik-Lehrbuch enthält. Mit Hilfe dieser Tafelwerte können, wenn sichergestellt ist, daß das statistische Material normalverteilt ist, praktische Fragen beantwortet werden, wie z. B.:

❶ Wieviel Prozent der männlichen Neugeborenen haben ein Körpergewicht zwischen 3400 und 3600 g?

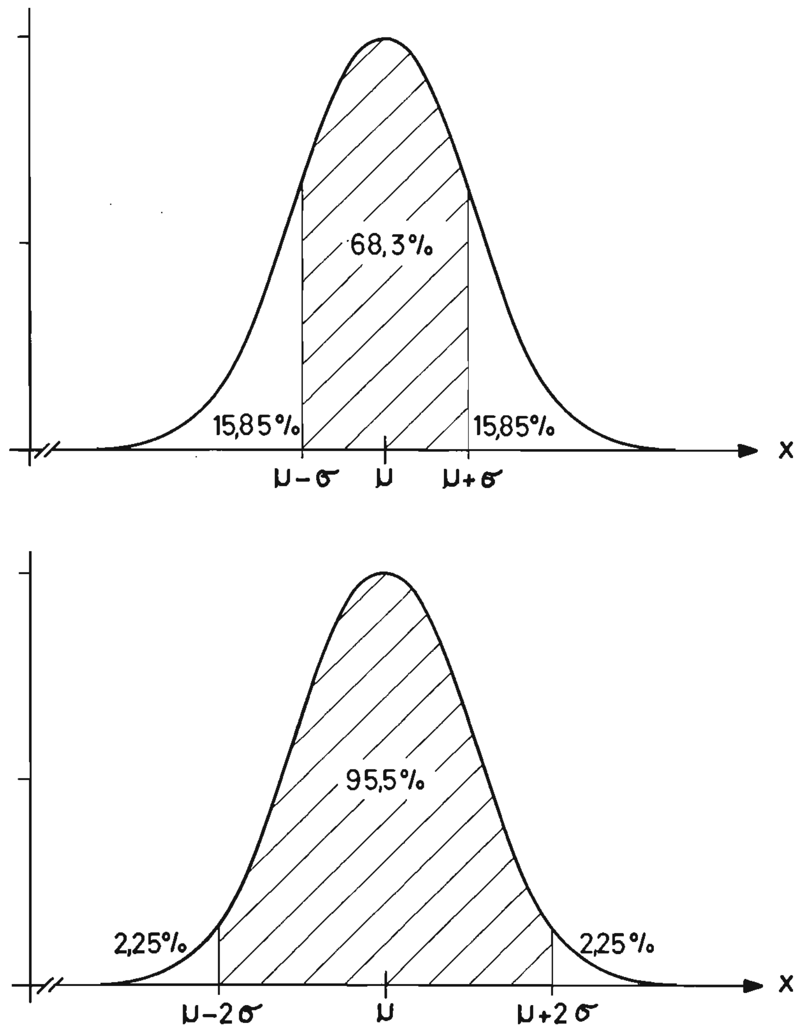
❷ Welche Grenzen hinsichtlich des Geburtsgewichtes müssen als Normbereich festgelegt werden, wenn 95 Prozent aller Ereignisse um den Mittelwert eingeschlossen werden sollen?

❸ Mit welcher Wahrscheinlichkeit liegt ein zufällig herausgegriffener Wert im Normbereich oder in einem beliebig vorzugebenden Bereich?

Für die Beantwortung dieser und ähnlicher Fragen ist entscheidend, daß die Normalverteilung so konstruiert ist, daß die über einem Bereich der x-Achse stehende Fläche immer die Wahrscheinlichkeit angibt, daß ein zufällig gewonnener Wert in diesen Bereich fällt. Zum Beispiel beträgt die Fläche über der Strecke

$\mu - \sigma$ bis $\mu + \sigma$: 68% der Gesamtfläche, die Fläche über
$\mu - 2\sigma$ bis $\mu + 2\sigma$: 95% der Gesamtfläche.

Oder umgekehrt: um 99 Prozent der Gesamtfläche im Zentrum zu erfassen, muß auf der x-Achse die Strecke von $\mu - 2{,}58\ \sigma$ bis $\mu + 2{,}58\ \sigma$ festgelegt werden. (Darstellung 2)



Darstellung 2: Der Flächeninhalt zwischen Glockenkurve, x-Achse und zwei auf der x-Achse errichteten Polen entspricht immer der Wahrscheinlichkeit, mit der ein beliebiger Wert in das auf der x-Achse abgesteckte Intervall fällt

In der Praxis sind statistische Größen dann normalverteilt, wenn viele zufällige Faktoren die Größe beeinflussen und diese weitgehend unabhängig voneinander sind. Diese Voraussetzung ist bei der biologischen Variabilität vieler Merkmale gegeben und daraus resultiert die außerordentliche Bedeutung der Normalverteilung in der medizinisch-biologischen Statistik. Kann man einmal eine experimentelle Gesamtheit beziehungsweise eine empirische Verteilung genügend genau durch die theoretisch-mathematische Normalverteilung beschreiben, kann für weitergehende Rechnungen und statistische Schlüsse die mathematische Verteilung benutzt werden, die als Voraussetzung aber die Normalverteilung haben.

Für die Prüfung, ob eine empirisch vorgelegte Verteilung durch eine Normalverteilung genügend genau angenähert werden kann, stehen statistische Prüfverfahren zur Verfügung.

Wahrscheinlichkeitspapier

Die Darstellung einer Normalverteilung auf dem sogenannten Wahrscheinlichkeitspapier liefert eine Gerade, da bei linearer Teilung der Abszisse die Ordinate entsprechend der Verteilungsfunktion der Normalverteilung geteilt ist. Diese Darstellung läßt eine anschauliche Beurteilung zu, wie weit die empirische Verteilung von der Normalverteilung abweicht. Ebenso lassen sich daraus Schätzwerte für den Mittelwert und die Varianz der empirischen Daten entnehmen. A. Habermehl

Literatur

Ramm, B.: Biomathematik, Ferdinand Enke Verlag, Stuttgart – Gross, R., u. Wichmann, H. E.: Was ist eigentlich normal?; Med. Welt 1979, 2–14.