Eine wesentliche Aufgabe der mathematischen Statistik ist es, die **Parameter unbekannter Verteilungen** zu ermitteln. Da dies in der Regel nur mittels der Auswertung von **Stichprobendaten** erfolgen kann, spricht man von **Parameterschätzung**. Geht man davon aus, dass die Parameter einer Verteilung einen zwar unbekannten, prinzipiell aber **eindeutigen Wert** haben, so können ihre auf Grund von Stichproben **geschätzten (empirischen) Werte** selbst als **Zufallsvariable mit einer bestimmten Verteilungsfunktion** aufgefasst werden.

Als Maß für die **Zuverlässigkeit eines Parameters** der Grundgesamtheit, den man auf Grund einer Stichprobe vom Umfang n **geschätzt** hat, dient die Angabe des **Konfidenzintervalls** (Vertrauensintervalls), in dem der richtige Wert des Parameters mit einer bestimmten, anzugebenden **Vertrauenswahrscheinlichkeit** P liegt.

Hier wird nur der in der Technik besonders wichtige Fall einer **normalverteilten Größe** X betrachtet, über deren **Mittelwert** durch Auswertung von Messungen eine seriöse Aussage gemacht werden soll, z.B. „$U = (1{,}75 \pm 0{,}05)\ \mathrm{V}$ mit $P = 95\%$“.

Steht eine **große Stichprobe** (Anhaltswert $n \geq 30$) zur Bestimmung der unbekannten Parameter **Mittelwert** $E(X) = \mu_x$ und **Standardabweichung** $\sqrt{\mathrm{Var}(X)} = \sigma_x$ zur Verfügung, so können die **Parameter**

empirischer Mittelwert $\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$, (1)

empirische Varianz $\boxed{s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}\left(x_i^2\right) - n\bar{x}^2\right]}$ bzw. (2)

empirische Standardabweichung[1] $\boxed{s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n}\left(x_i^2\right) - n\bar{x}^2\right]}}$ (3)

der Stichprobe direkt als Schätzwerte für die unbekannten Parameter verwendet werden.
Der berechnete Mittelwert $\bar{x}$ ist selbst eine Zufallsvariable, die **normalverteilt** ist mit der

Varianz des Mittelwertes $\boxed{s_{\bar{x}}^2 = \frac{s_x^2}{n}}$ bzw. der (4)

Standardabweichung des Mittelwertes $\boxed{s_{\bar{x}} = \frac{s_x}{\sqrt{n}}}$. (5)

Somit lässt sich aus einer Wertetabelle für die **Standard-Normalverteilung** $\Phi(z)$ die Wahrscheinlichkeit P, dass der gesuchte Wert μ_x im Bereich $\bar{x} - t \cdot s_{\bar{x}} \leq \mu_x \leq \bar{x} + t \cdot s_{\bar{x}}$ liegt, näherungsweise berechnen als $\boxed{P\left(\left|\bar{x} - \mu_x\right| \leq t \cdot s_{\bar{x}}\right) = 2\Phi(t) - 1}$. (6)

Ist die **Stichprobe klein** (Anhaltswert $n < 30$), so werden die aus den wenigen Messwerten mittels (1) und (3) berechneten empirischen Parameter mit sinkendem n drastisch unsicherer und man muss an Stelle der Standard-Normalverteilung die **t-Verteilung**[2], verwenden.

---

[1] Diese Funktion wird auf vielen Taschenrechnern (u.a. auf dem TI 30) mit $\sigma_{x\,n-1}$ bezeichnet.

[2] Diese auch Student-t-Verteilung genannte Verteilung wurde im Jahre 1908 von dem englischen Mathematiker William S. Gosset (1876-1937) unter dem Pseudonym „Student“ veröffentlicht.

Während der Wert der Standard-Normalverteilung $\Phi(z)$ nur von <u>einer</u> unabhängigen Variablen $z = \frac{x - \mu_x}{\sigma_x}$ abhängt, ist der Wert der Verteilungsfunktion $F_t(t,k)$ der t-Verteilung $\boxed{F_t(t,k) = \int_{-\infty}^{t} f_t(u,k)\,du}$ mit der Dichtefunktion $\boxed{f_t(t,k) = A_k (1 + t^2/k)^{-\frac{k+1}{2}}}$ von einer unabhängigen Variablen $t = \frac{\bar{x} - \mu_x}{s_{\bar{x}}}$ mit $\bar{x}$ nach (1) und $s_{\bar{x}}$ nach (3) <u>und zusätzlich</u> einem **Parameter k** mit $k \geq 1$, der **Anzahl der Freiheitsgrade** der Verteilung, abhängig. Somit handelt es sich bei „der" t-Verteilung eigentlich um eine ganze **Klasse von Verteilungen**: je eine für jeden Wert von k. Je kleiner k ist, desto flacher verläuft die Funktion. Mit steigendem Wert von k nähert sich der Verlauf der t-Verteilung immer mehr dem Verlauf der Standard-Normalverteilung an.

Die Dichtefunktion der t-Verteilung ist eine gerade Funktion. Ihre Kennwerte sind $\mu = 0$ für $k \geq 2$ und $\sigma^2 = \frac{k}{k-2}$ für $k \geq 3$. Der Wert des vom Parameter k abhängigen Faktors A_k in der Dichtefunktion ergibt sich aus der Normierungsbedingung $\int_{-\infty}^{\infty} f_t(t,k) = 1$.

Ebenso wie die Funktionswerte der Standard-Normalverteilung erhält man die **Funktionswerte** der t-Verteilung **aus Tabellen in der Fachliteratur**. Die Werte der t-Verteilung sind aber <u>nicht</u> als Funktion von t tabelliert. Vielmehr werden diejenigen Werte von t angegeben, die zu einem bestimmten **Parameterpaar** gehören, nämlich dem **Freiheitsgrad** $\boxed{k = n - 1}$ und einem bestimmten **Wahrscheinlichkeitswert**. Diese Wahrscheinlichkeit ist entweder die **Vertrauenswahrscheinlichkeit** P, mit der die Zufallsvariable im Bereich $-t..+t$ liegt (z.B. 90%, 95%, 98%, 99%, 99,9%) oder oftmals auch die **halbe Irrtumswahrscheinlichkeit** $\alpha = 0{,}5(1 - P)$. Es gibt noch diverse andere Darstellungsvarianten.

Tabelle 1 zeigt eine kleine derartige Tabelle, in der als Parameter die Vertrauenswahrscheinlichkeit P verwendet wird. Für $k \to \infty$ gilt $F_t(k,t) = \Phi(t)$.

Tabelle 1: Ausgewählte Werte des Parameters t

k	P		
	95%	99%	99,9%
1	12,706	63,657	636,619
2	4,303	9,925	31,598
3	3,182	5,841	12,924
4	2,776	4,604	8,610
5	2,571	4,032	6,869
6	2.447	3,707	5,959
7	2,365	3,499	5,408
8	2,306	3,355	5,041
9	2,262	3,250	4,781
10	2,228	3,169	4,587
15	2,131	2,947	4,073
20	2,086	2,845	3,890
25	2,060	2,787	3,725
30	2,042	2,750	3,646
40	2,021	2,704	3,551
50	2,010	2,678	3,496
∞	1,960	2,576	3,291