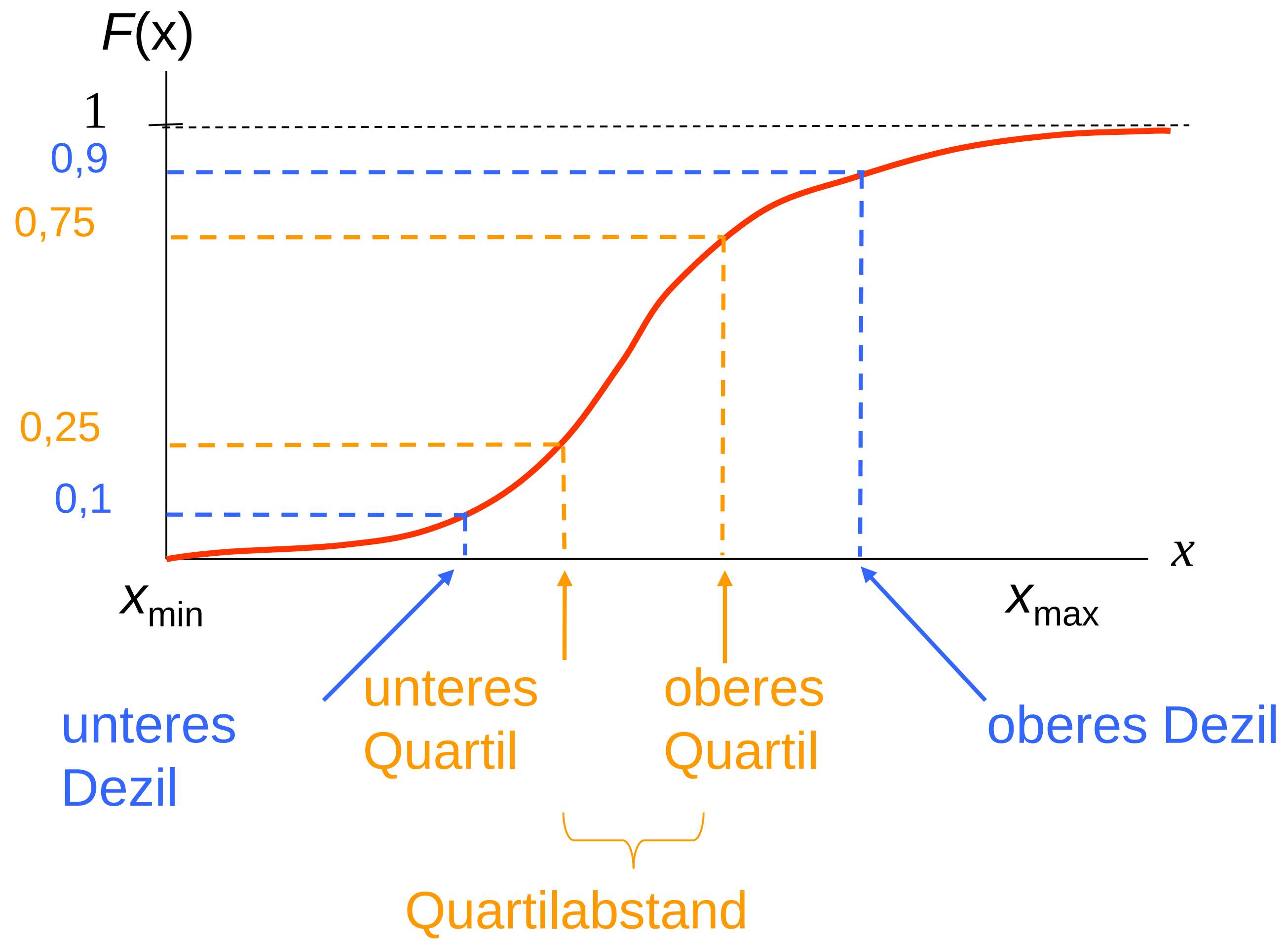
Quantile und die relative Summenhäufigkeits-verteilung

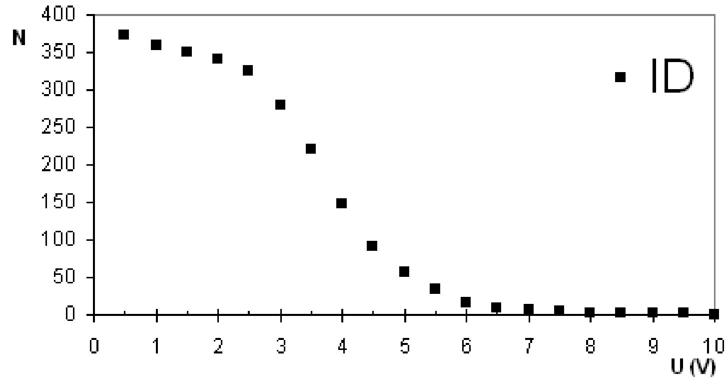


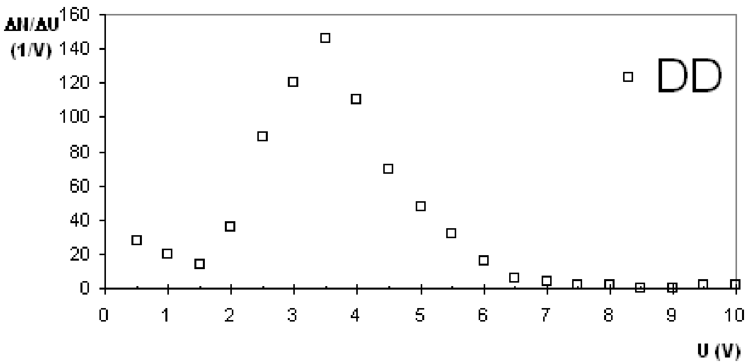
Beispiel in der Physikpraktikum:

Coulter Zähler

(siehe viel später...)

	A	B	C
1	**Coulter counter**		
2		ID	DD
3		N	ΔN/ΔU
4	V		1/V
5	0.5	373	28
6	1	359	20
7	1.5	349	14
8	2	342	36
9	2.5	324	88
10	3	280	120
11	3.5	220	146
12	4	147	110
13	4.5	92	70
14	5	57	48
15	5.5	33	32
16	6	17	16
17	6.5	9	6
18	7	6	4
19	7.5	4	2
20	8	3	2
21	8.5	2	0
22	9	2	0
23	9.5	2	2
24	10	1	2





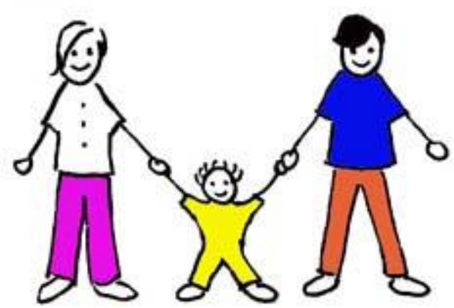
Verteilungen und Schätzungen
Grundlagen der Wahrscheinlichkeitslehre

Zufallsexperiment

- Vorgang nach einer bestimmten Vorschrift ausgeführt
- (im Prinzip) beliebig oft wiederholbar
- sein Ergebnis ist zufallsabhängig (in der Natur ist es immer!) Es gibt eine eingebaute Unsicherheit in der Natur.
- bei mehrmaligen Durchführung des Experiments beeinflussen die Ergebnisse einander nicht

Würfelspiel

Roulett



$I^A i$? $I^A i$

Blutgruppenversuch

Elementarereignisse

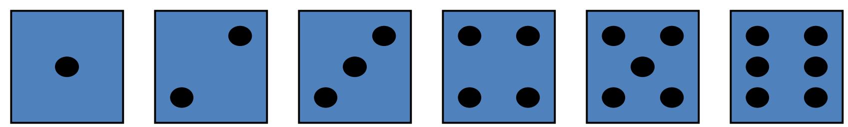
die einzelnen, nicht mehr zerlegbaren und sich gegenseitig ausschliessenden Ausgänge oder Ergebnisse eines Zufallsexperimentes

Ereignismenge, Ereignisraum (Ω)

Reihe aller möglichen Elementarereignisse. Z.B:

beim Würfelspiel:

$\Omega = \{1, 2, 3, 4, 5, 6\}$



beim Münzenexperiment: $\Omega = \{\text{Zahl}, \text{Kopf}\}$

beim „Blutgruppenversuch”: $\Omega = \{I^A I^A,\ I^A i,\ iI^A,\ ii\}$

Wahrscheinlichkeit

Bernoulli (1654-1705), Laplace (1749-1827)
(**klassische Wahrscheinlichkeit**)

Bei einem Zufallsexperiment, was endlich viele Ausgänge hat, die (zB. wegen Symmetriegründen) **gleichwahrscheinlich** sind, die Wahrscheinlichkeit eines Ereignisses (*E*) ist:

$$\boxed{p(E) = \frac{\text{Anzahl der für } E \text{ günstigen Elementarereignisse}}{\text{Anzahl aller gleichmöglichen Elementarereignisse}}}$$

Dabei denken wir, das alle interessante Ereignisse eigentlich aus Kombinationen verschiedener Elementarereignisse aufbaubar sind, der Anzahl wovon kann auch sehr gross sein (wie im Lego-Spiel).

p=probability, Probabilität

Beispiele

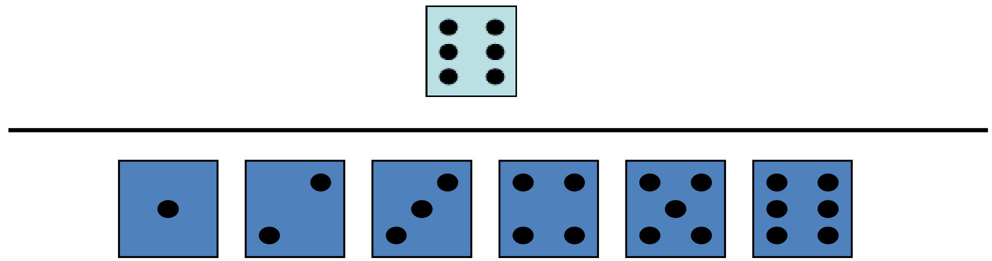
$$\boxed{p(E) = \frac{g}{m}}$$

günstig

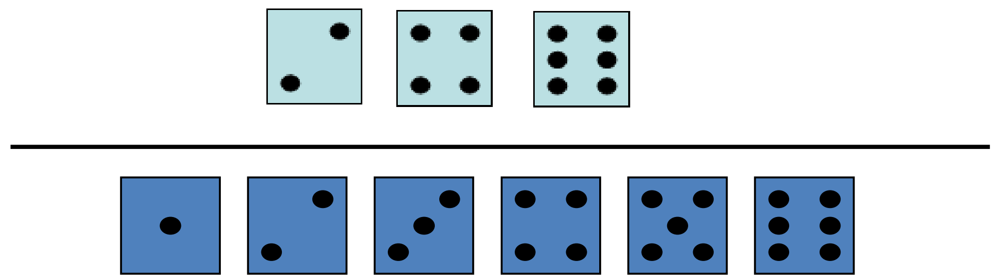
alle

Würfelexperiment:

$$p(6) = \frac{1}{6}$$



$$p(gerade\ Zahl) = \frac{3}{6}$$



Münzenexperiment:

$$p(Kopf) = \frac{1}{2}$$

Statistische Wahrscheinlichkeit:
oft sind die Elementarereignisse NICHT gleich wahrscheinlich!

Zufallsexperiment $\longrightarrow$

Ereignis *A*

Ereignis *B*

Gefälschte Münze?

Tritt bei *n*-maliger Durchführung eines Zufallsexperimentes ein bestimmtes Ereignis ***A*** *k*-mal auf, so bezeichnet man die in langen Versuchsreihen zu beobachtende relative Häufigkeit als **Wahrscheinlichkeit**, ***p(A)*** :

$$\boxed{p(A) = \frac{k}{n}}$$

Wenn n -> unendlich

Eigenschaften der Wahrscheinlichkeit

- $0 \leq p(A) \leq 1$
- p(sicheres Ereignis) = 1
- p(unmögliches Ereignis) = 0

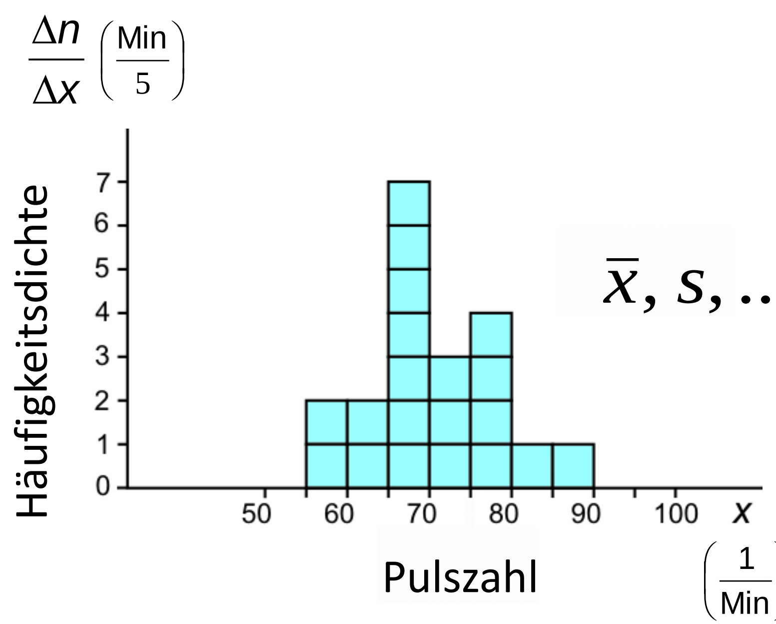
Verteilungen

Population

Wahrscheinlichkeitsverteilung

?, ?, ...

Stichprobe



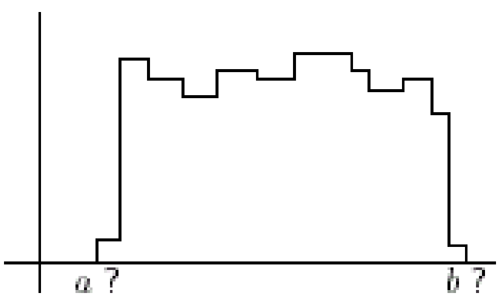
$\bar{x}, s, \ldots$

Verteilungen

Wie kann man die theoretische Verteilung bestimmen?

Vermutung

(nach dem Histogram)



Gleichverteilung?

Modellannahme

Laplace-Prinzip:
wenn nichts dagegen spricht, gehen wir davon aus, dass alle Elementarereignisse gleich wahrscheinlich sind

Laplace-Experiment:
es meint ein Zufalls-Experiment bei dem davon ausgegangen wird, dass jeder Versuchsausgang **gleichwahrscheinlich** ist

Gleichverteilung der Elementarereignisse

Klassifizierung der Verteilungen

- **diskrete Verteilungen**
- diskrete Gleichverteilung
- Binomialverteilung
- Poisson Verteilung
- ...

diskrete Zufallsgröße

zB: Anzahl der Kranken, Augenzahl des Würfels

- **kontinuierliche Verteilungen**
- kontinuierliche Gleichverteilung
- Normalverteilung
- Chi-Quadrat Verteilung
- *t*-Verteilung
- ...

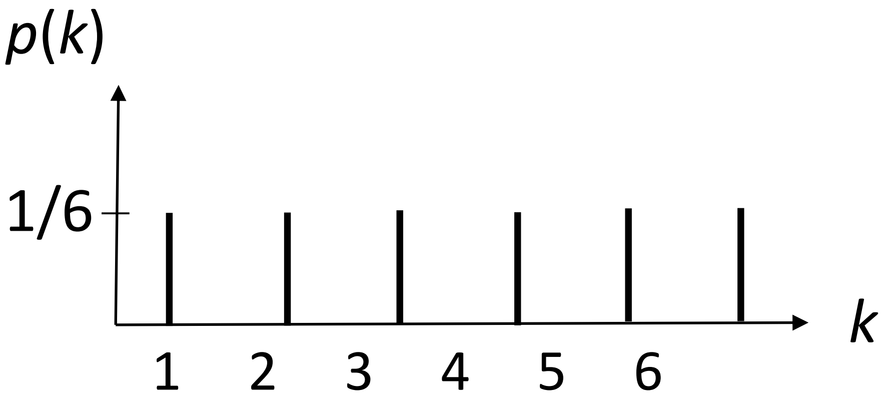
kontinuierliche Zufallsgröße

zB: Blutdruck, Körperhöhe,...

Diskrete Gleichverteilung

Beispiel:

Wertebereich	1	2	3	4	5	6
Wahrscheinlichkeit	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



weitere Beispiele:

Münzenversuch

Würfelexperiment
mit einem Ikosaeder

$$p(k) = \frac{1}{6}, \quad k = 1, 2, \ldots, 6$$

Lageparameter der Verteilung

Es sei X eine diskrete Zufallsgröße mit Werten $x_1, x_2 \ldots$ dann heisst

$$\boxed{\mu = \sum_i x_i p(x_i)}$$

Erwartungswert von *X.*

Der Erwartungswert gibt denjenigen Wert an, den man als Mittelwert (durchschnittlichen Wert) über viele Versuchswiederholungen "erwarten" kann.

Dabei ist es durchaus möglich, dass der Erwartungswert bei keinem einzigen Versuch realisiert wird oder sogar überhaupt nicht vorkommen kann.

Erwartungswert und Durchschnittswert

$$\mu = \sum_i x_i p(x_i) \qquad \bar{x} = \sum_i x_i h_i$$

Rel.Häufigkeit

Beispiel: 100 Würfelexperimente. 2,5,4,3,6,6,1,5,4,2,3...
Insgesamt:

x_i	n_i	h_i
1	15	15/100
2	20	20/100
3	14	14/100
4	16	16/100
5	18	18/100
6	17	17/100

$$\bar{x} = \frac{15\cdot 1 + 20\cdot 2 + 14\cdot 3 + 16\cdot 4 + 18\cdot 5 + 17\cdot 6}{100} =$$

$$= \frac{15}{100}\cdot 1 + \frac{20}{100}\cdot 2 + \frac{14}{100}\cdot 3 + \frac{16}{100}\cdot 4 + \frac{18}{100}\cdot 5 + \frac{17}{100}\cdot 6 = 3.53 =$$

$$= h(1)\cdot 1 + h(2)\cdot 2 + h(3)\cdot 3 + h(4)\cdot 4 + h(5)\cdot 5 + h(6)\cdot 6 \to$$

$$\xrightarrow[n\to\infty]{} P(1)\cdot 1 + P(2)\cdot 2 + P(3)\cdot 3 + P(4)\cdot 4 + P(5)\cdot 5 + P(6)\cdot 6 = \mu$$

x_i: Augenzahl

n_i: absolute Häufigkeit

h_i: relative Häufigkeit

$$\boxed{\bar{x} \xrightarrow[n\to\infty]{} \mu}$$

Streuung der Verteilung

Es sei *X* eine diskrete Zufallsgröße mit Werten $x_1, x_2, \ldots$ und mit dem Erwartungswert μ. Dann nennt man die Zahl

$$\boxed{\sigma^2 = \sum_i (x_i - \mu)^2 p(x_i)}$$

als Varianz von *X*, ihre Wurzel als (theoretische) Streuung (σ).

$$s \xrightarrow[n \to \infty]{} \sigma$$

empirische Streuung $\rightarrow$ theoretische Streuung

(Standardabweichung)

Normalverteilung

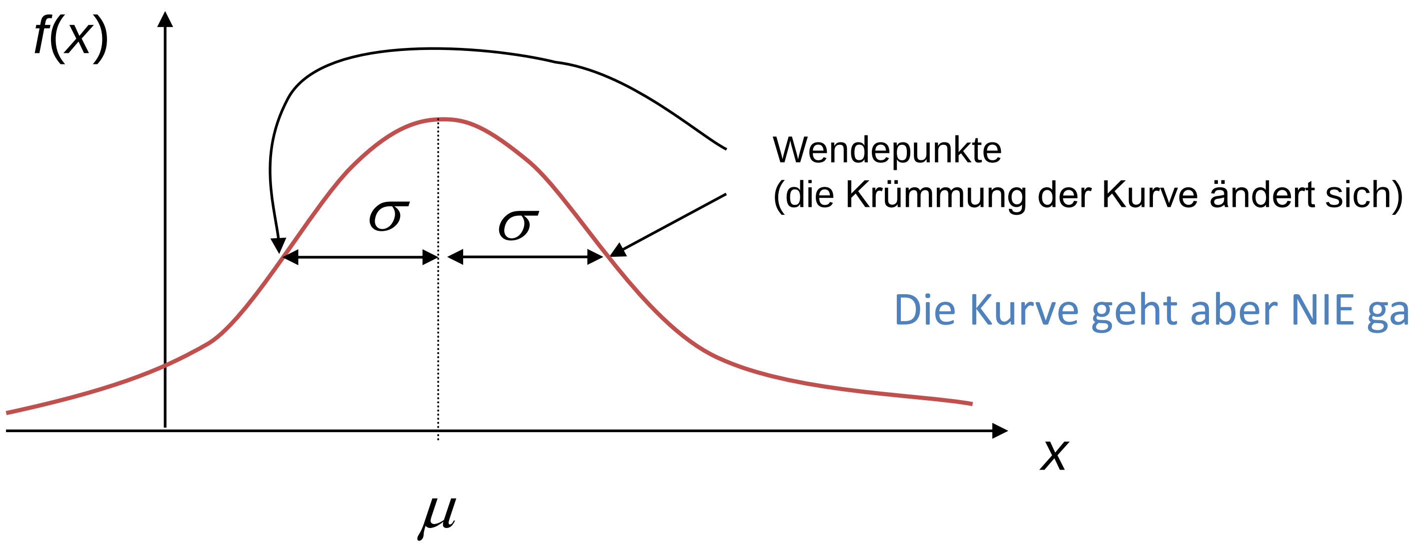
Verteilungsdichtefunktion:

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

Parameter der Normalverteilung:

Erwartungswert: μ
Streuung: σ

Oberfläche unter der Kurve = 1.
(gilt für alle verteilungsdichtefuntionen!)



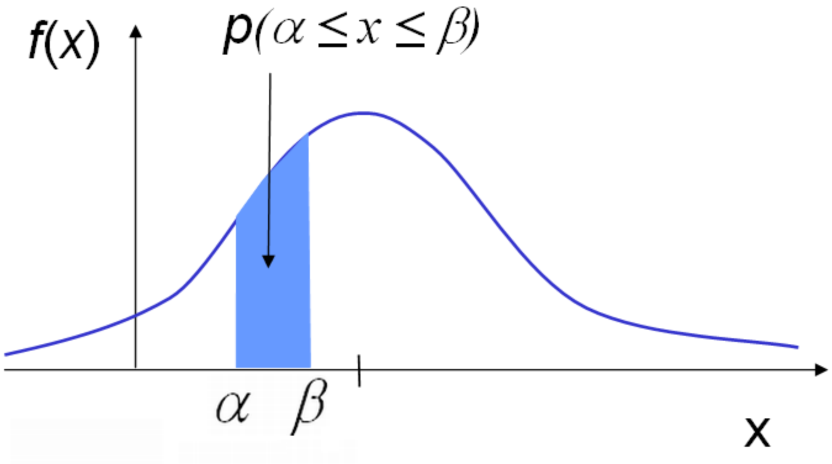
Die Kurve geht aber NIE ganz zu Null!

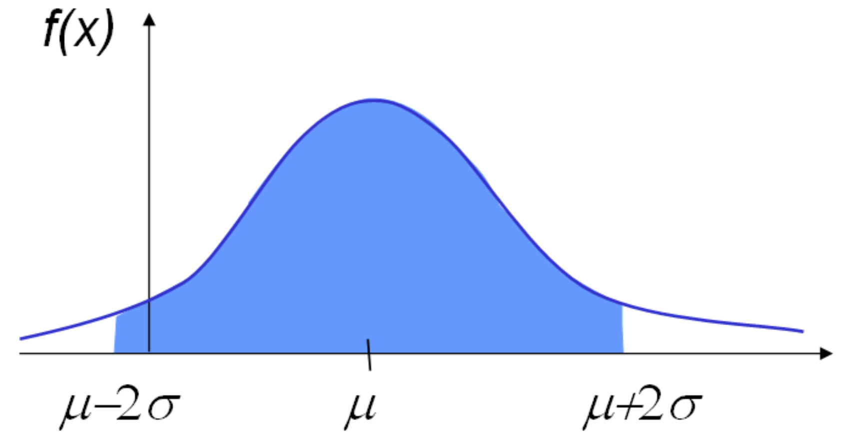
Normalverteilung (Gauss-Verteilung)

für die dargestellte Funktion: $\mu = 3,\ \sigma = 1$

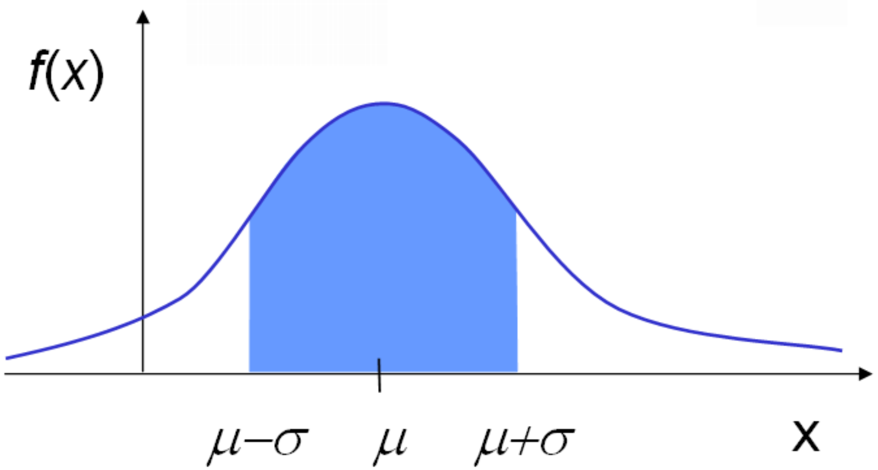
Normalverteilung

Wahrscheinlichkeit ist eine Oberfläche unter der Dichtefunktion!

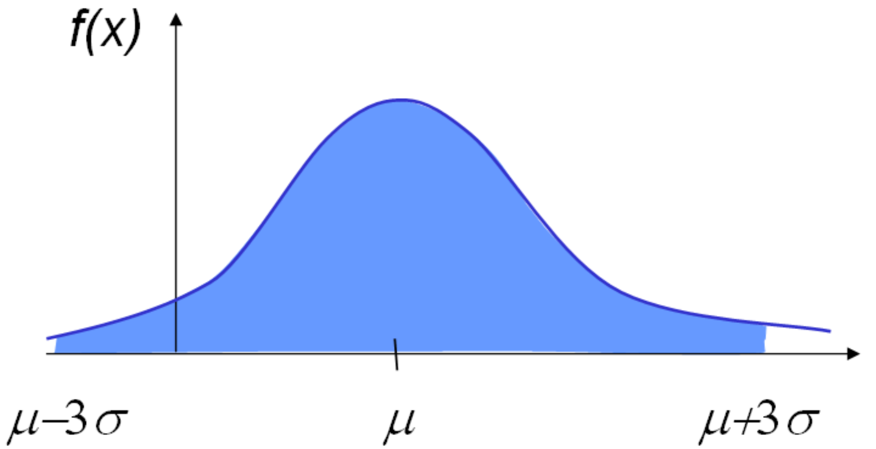




$p(\mu - 2\sigma \le x \le \mu + 2\sigma) = 95\%$



$p(\mu - \sigma \le x \le \mu + \sigma) = 68\%$



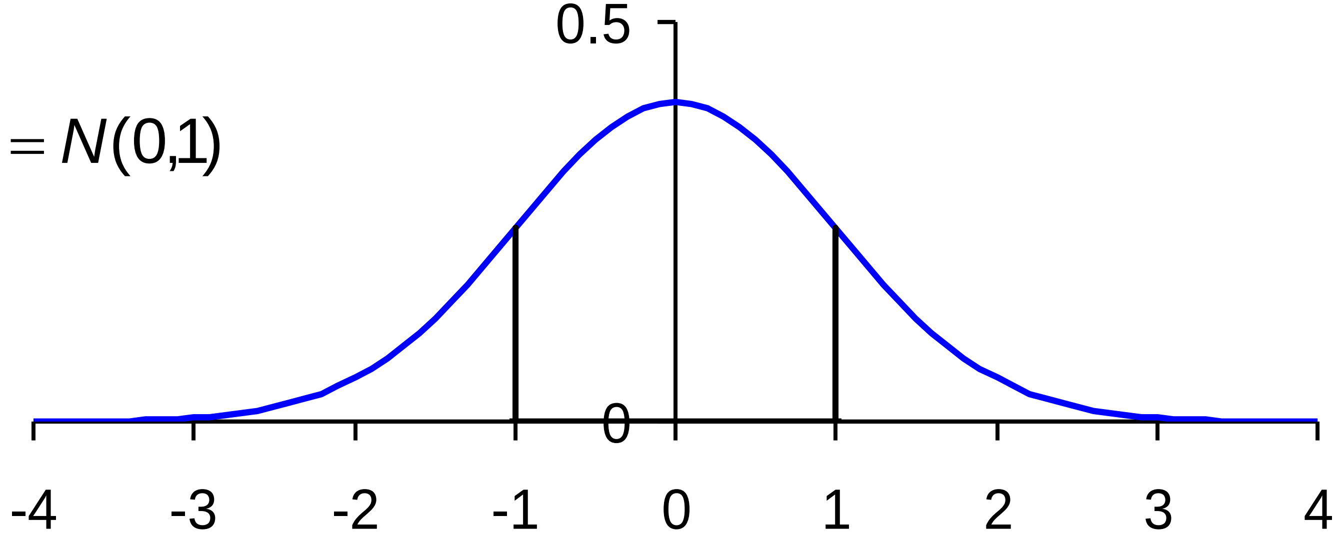
$p(\mu - 3\sigma \le x \le \mu + 3\sigma) = 99{,}8\%$

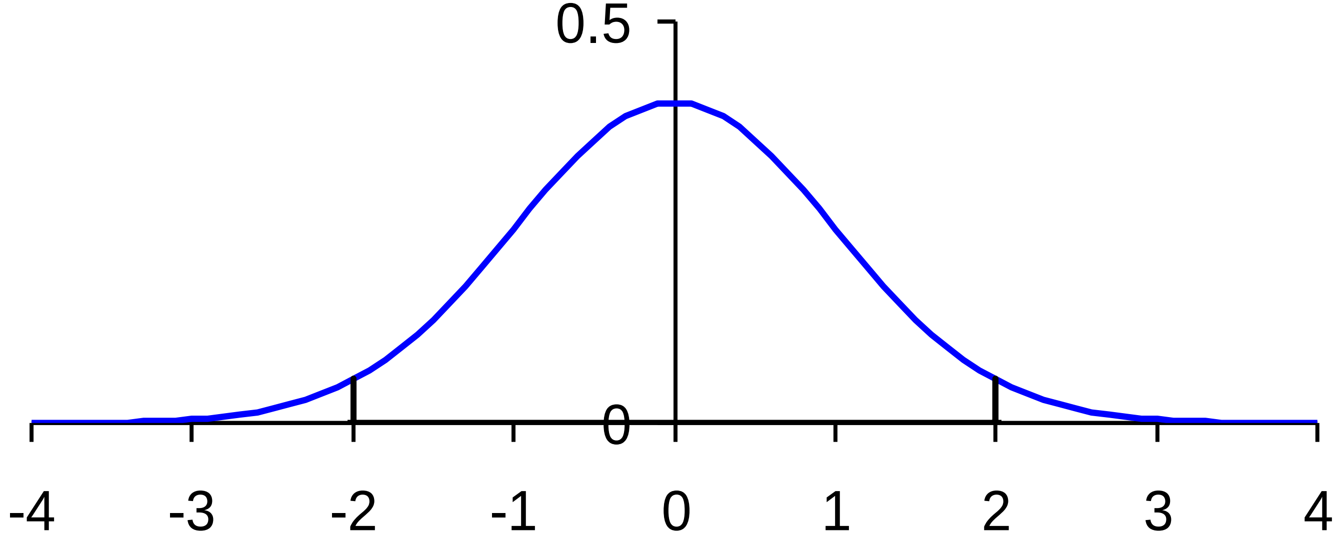
Standard - Normalverteilung

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = N(0,1)$$

$$\mu = 0$$

$$\sigma = 1$$



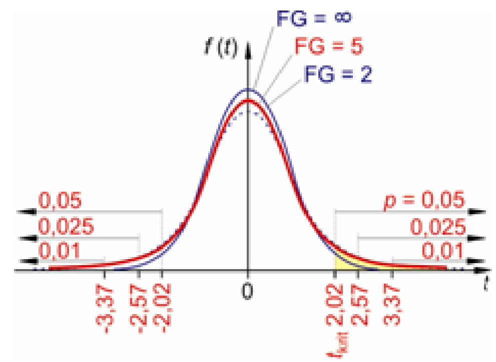


1. STATISTISCHE TABELLEN

t-Verteilungsfamilie

t-VERTEILUNG

	p (Irrtumswahrscheinlichkeit, einseitiger Test)						
Freiheitsgrad (FG)	**0,4**	**0,25**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**
	p (Irrtumswahrscheinlichkeit, zweiseitiger Test)						
	0,8	**0,5**	**0,2**	**0,1**	**0,05**	**0,02**	**0,01**
1	0,325	1,000	3,078	6,314	12,70	31,82	63,65
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499
25	0,256	0,684	1,316	1,708	2,060	2,485	2,787
26	0,256	0,684	1,315	1,706	2,056	2,479	2,779
27	0,256	0,684	1,314	1,703	2,052	2,473	2,771
28	0,256	0,683	1,313	1,701	2,048	2,467	2,763
29	0,256	0,683	1,311	1,699	2,045	2,462	2,756
30	0,256	0,683	1,310	1,697	2,042	2,457	2,750
40	0,255	0,681	1,303	1,684	2,021	2,423	2,704
60	0,255	0,679	1,296	1,671	2,000	2,390	2,66
120	0,254	0,677	1,289	1,658	1,980	2,358	2,617
∞	0,250	0,674	1,282	1,645	1,960	2,326	2,576



„Glockenkurven”

Je größer ist der Freiheitsgrad, desto schmaler ist die Kurve.

Also der Freiheitsgrad ist ein Zahl um eine bestimmte Kurve auszuwählen.

$$\boxed{t_\infty \equiv N(0,1)}$$

Pr.Buch Anhang S.27.2

Zentraler Grenzwertsatz

•Es seien x_1, $x_{2,\ldots}x_n$ unabhängige Zufallsgrößen, die alle derselben Verteilung haben.

•Die **Verteilung der Summe** nähert sich einer **Normalverteilung,** wenn $n \to \infty$. $S_n = \sum_{i=1}^{n} x_i$

•Die Summe der Verteilungsfunktionen konvergiert gegen eine Normalverteilung auch wenn die einzelnen Zufallsgrößen keine Normalverteilung haben.

•**Biologische Bedeutung:**
Wenn ein Parameter (zB. Körpergröße, Blutzuckerkonzentration) durch viele anderen Faktoren (Zufallsgrößen) beeinflusst wird, folgt dieser Parameter einer Normalverteilung.

Analytische Statistik

Population

N = „unendlich”

Stichprobe

n = endlich

Theoretische Verteilung
Erwartungswert
Theoretische Streuung

Häufigkeitsverteilung
Durchschnitt
Standardabweichung

Aufgabe der Schätztheorie

Aus einer Stichprobe Schätzwerte für

- Wahrscheinlichkeiten
- Erwartungswert
- Streuung
- oder andere Parametern

einer Verteilung zu ermitteln.

Typen der Schätzungen:

- ***Punktschätzung***
- ***Intervallschätzung***

Punktschätzungen

- Der Parameter wird mit einem Wert geschätzt.

- Relative Häufigkeit
  ist ein Schätzwert für die Wahrscheinlichkeit

- Durchschnitt
  ist ein Schätzwert für den Erwartungswert

- Standardabweichung
  ist ein Schätzwert für die Streuung

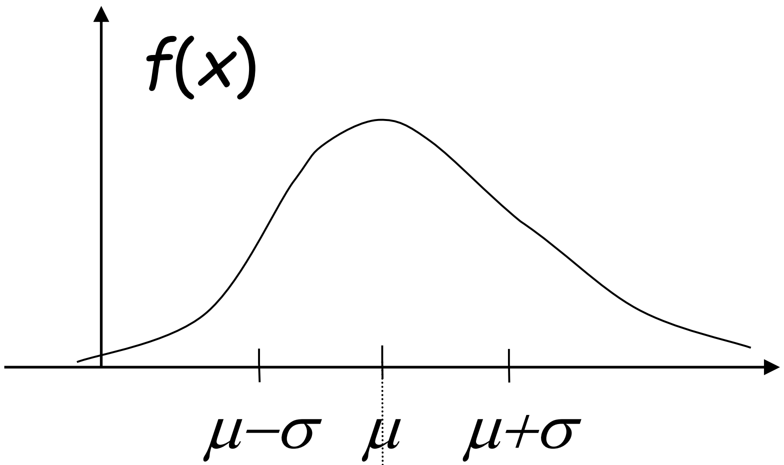
- Punktschätzungen sagen
  nichts über die Genauigkeit bzw. Sicherheit der Schätzung!

Intervallschätzungen

•Intervallschätzung oder Konfidenzschätzung gibt zu einer vorgewählten Sicherheitswahrscheinlichkeit γ, (Konfidenzniveau) ein Intervall (c_1,c_2) an, in dem der unbekannte Parameter (zB. μ oder σ) mit einer Wahrscheinlichkeit von mindestens γ liegt.

Zb.: Erwartungswert der Pulszahl ist bei
95% Konfidenzniveau: $74 \pm 6\ ^1/_{Min}$

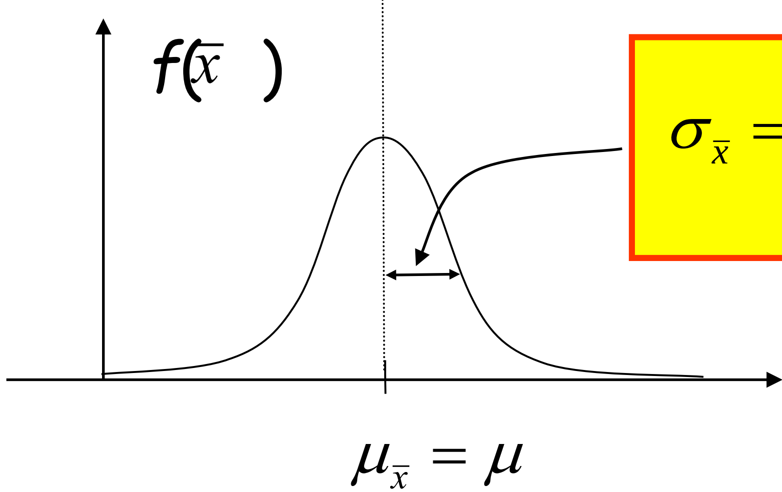
Konfidenzintervall für den Erwartungswert



x zB:Körperhöhe

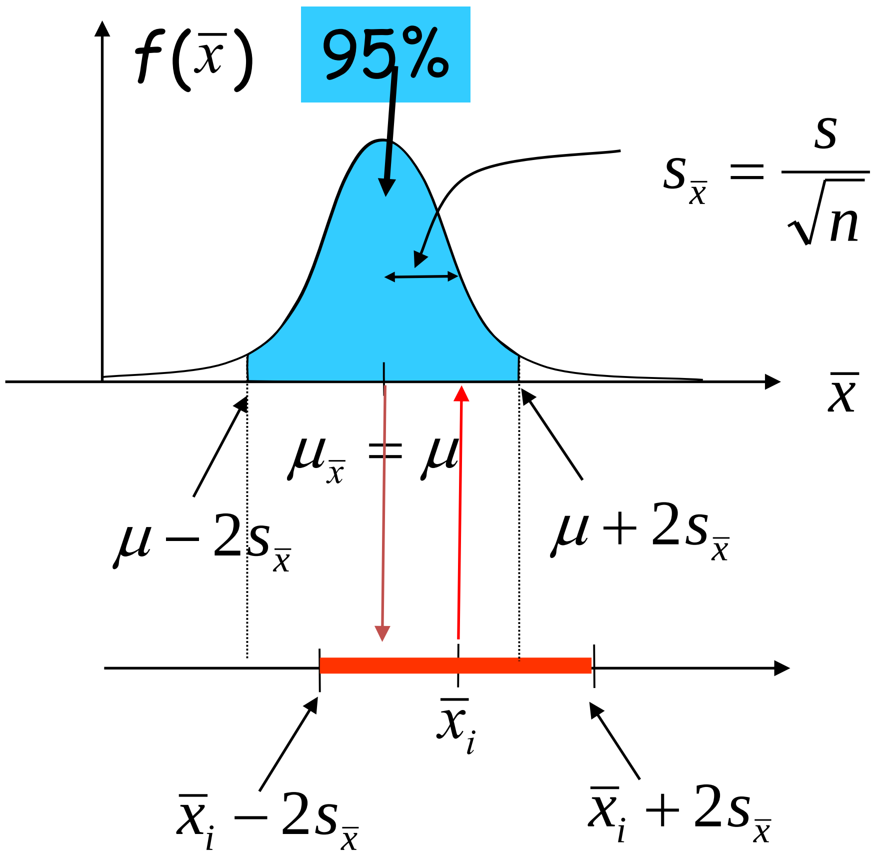
Standardfehler

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx s_{\bar{x}}$$



$\bar{x}$ zB: durchschnittliche Körperhöhe in einem Studentengruppe von *n* Studenten

Konfidenzintervall für den Erwartungswert



$\bar{x}_i$ liegt mit 95% Wahrscheinlichkeit im Intervall

$\mu - 2s_{\bar{x}} \quad \mu + 2s_{\bar{x}}$

Und gleichzeitig mit 100-95= 5% Wahrscheinlichkeit irgendwo draussen!

wenn $\mu - 2s_{\bar{x}} \leq \bar{x}_i \leq \mu + 2s_{\bar{x}}$ dann $\boxed{\bar{x}_i - 2s_{\bar{x}}} \leq \mu \leq \boxed{\bar{x}_i + 2s_{\bar{x}}}$

95% Wahrsch. 95% Wahrsch.

Konfidenzintervall für den Erwartungswert

In dem Intervall $\bar{x} - 2s_{\bar{x}},\ \bar{x} + 2s_{\bar{x}}$ **Konfidenzintervall**
liegt der Erwartungswert (μ) mit 95% Wahrscheinlichkeit

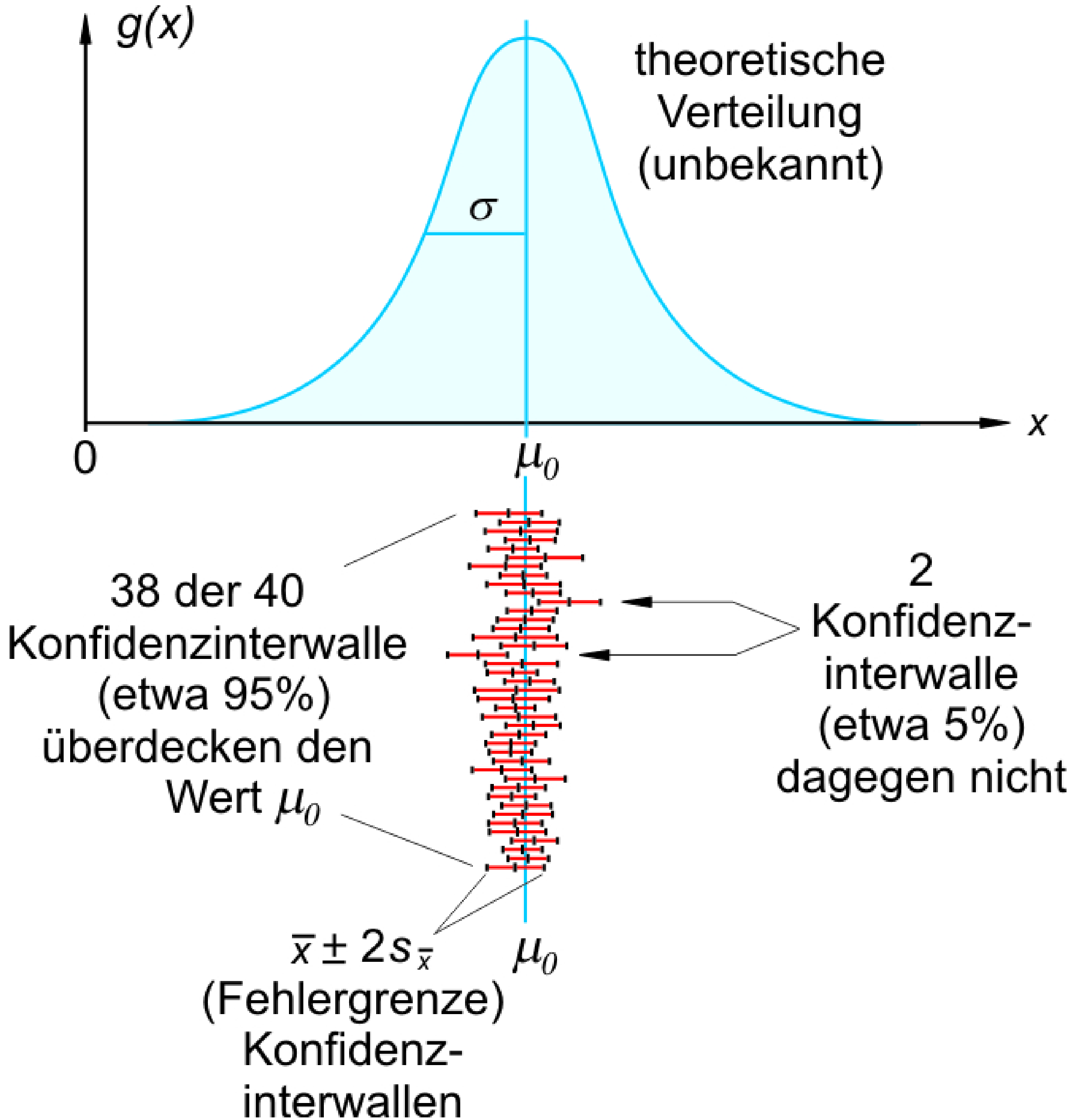
Eine ähnliche Ableitung gibt: μ liegt
-mit 68% Wahrscheinlichkeit im Intervall: $\bar{x} - s_{\bar{x}},\ \bar{x} + s_{\bar{x}}$

- mit 99,7% Wahrscheinlichkeit im Intervall:

$$\bar{x} - 3s_{\bar{x}},\ \bar{x} + 3s_{\bar{x}}$$

Je größer ist die Sicherheitswahrscheinlichkeit desto breiter ist das Konfidenzintervall!

Bemerkung: wenn n$\rightarrow\infty$ dann $s_{\bar{x}} \rightarrow 0$



Pr.Buch Abb. 11

Zusammenfassung der Schätzungen

- Punktsätzungen:

Stich-probe		Grund-gesamtheit
$\bar{x}$	$\longrightarrow$	μ
s	$\longrightarrow$	σ
n	$\longrightarrow$	∞

Intervallschätzung für den Erwartunswert:

$\boxed{\bar{x} \pm 2 s_{\bar{x}} \quad 95\%}$