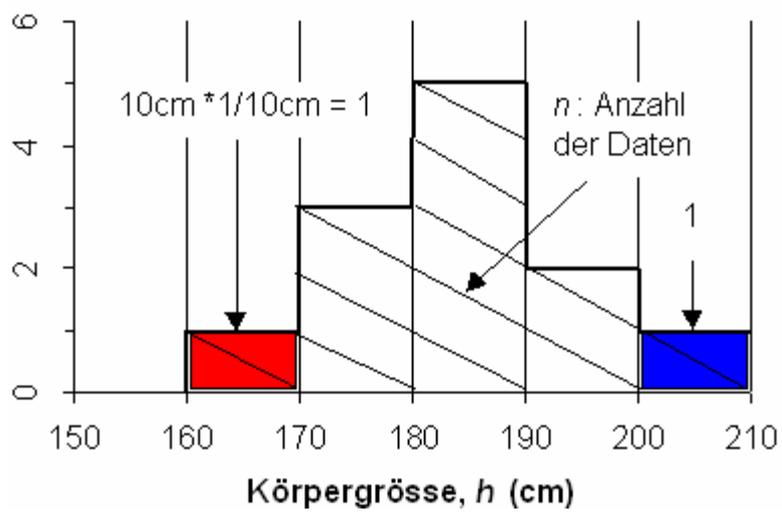


Häufigkeitsverteilung

$$\frac{\Delta N}{\Delta h}$$

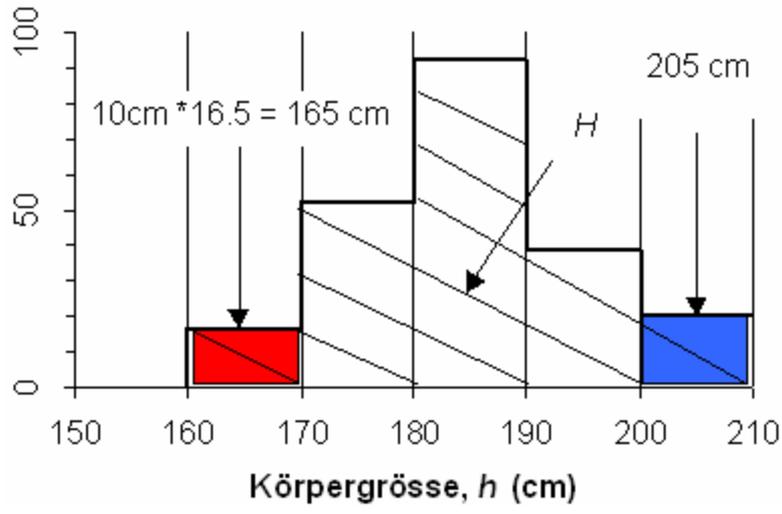
$$\left(\frac{1}{10\text{cm}} \right)$$



h : Körperhöhe

H : kollektive Höhe, Gesamthöhe

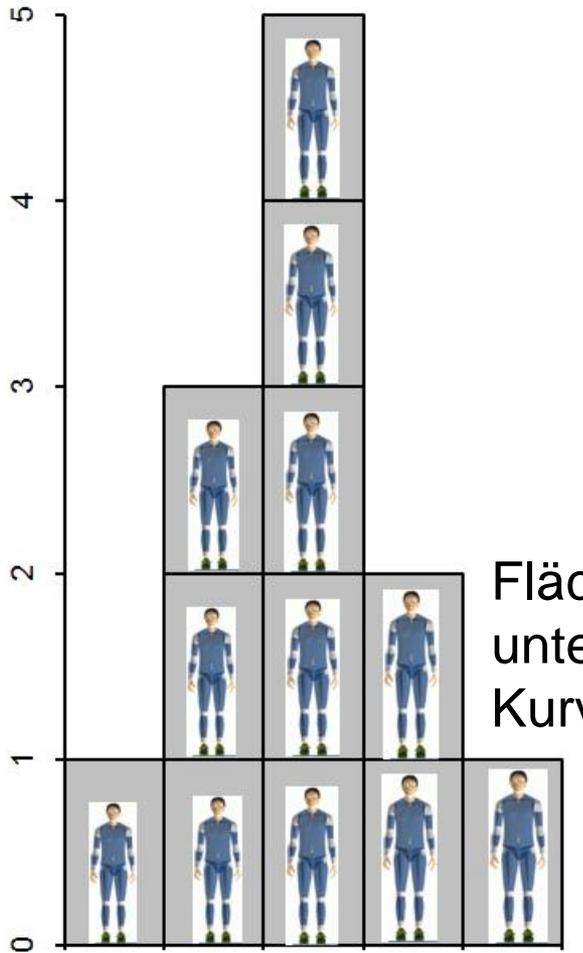
$$\frac{\Delta H}{\Delta h}$$



Spektrum ist eine spezielle Häufigkeitsverteilung

Häufigkeitsdichte

$$\frac{\Delta N}{\Delta h} \left(\frac{1}{10\text{cm}} \right)$$

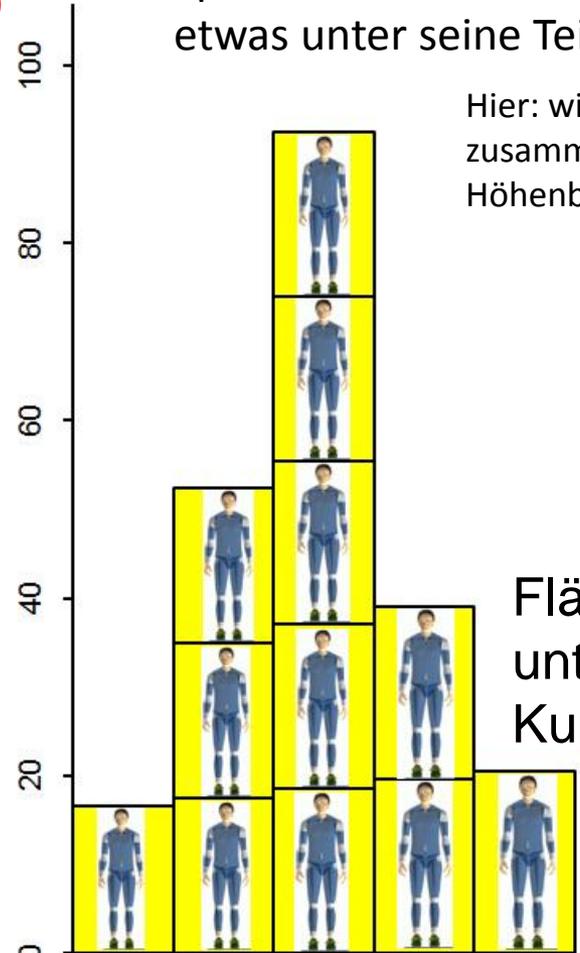


Fläche unter der Kurve: n

160 170 180 190 200 210

h (cm)

$$\frac{\Delta H}{\Delta h}$$



Spektrum ist die Verteilung von etwas unter seine Teile

Hier: wie viel Höhe ist zusammen in einem Höhenbereich

Fläche unter der Kurve: H

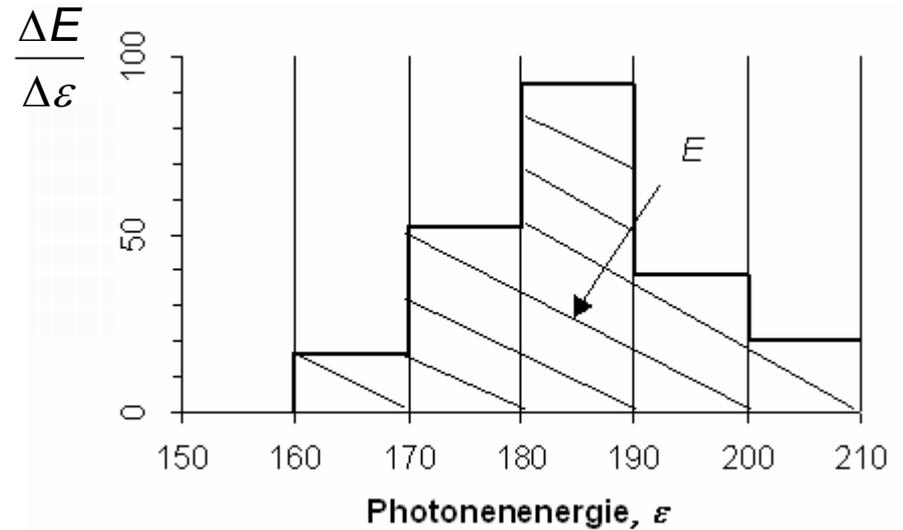
160 170 180 190 200 210

h (cm)

Beispiel aus der Physik

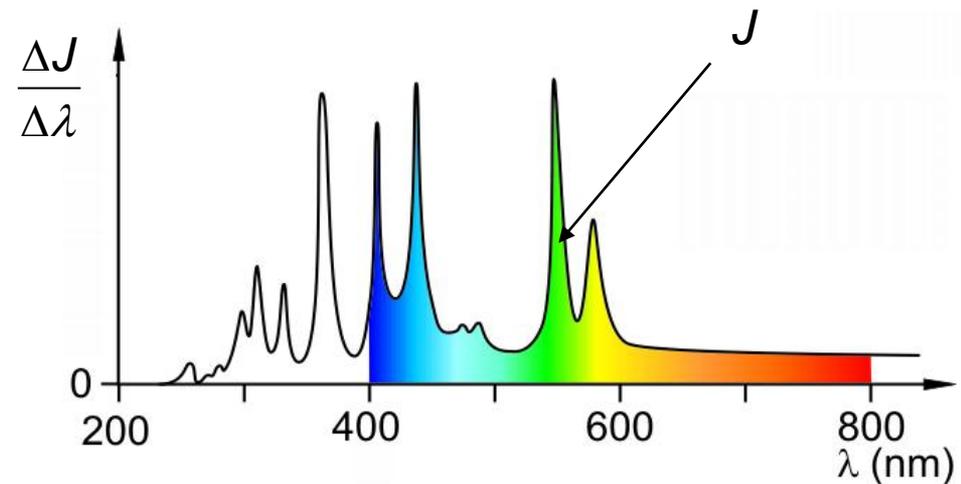
Emissionsspektrum:

wie verteilt sich die emittierte Energie über die Photonenenergien



Intensität: $J = \Delta E / (\Delta A \cdot \Delta t)$ [J/m^2s]

Benützung der Wellenlänge(λ) ist bequemer als die der Photonenenergie. $\lambda = c/f$



Lageparameter. Charakterisierung des Zentrums der Daten

Durchschnittswert (der arithmetische Mittelwert)

=average(...)
=Mittelwert(...)

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Modus (Modalwert, Dichtemittel): der Wert mit der größten Wahrscheinlichkeit;
der häufigste Wert einer Häufigkeitsverteilung

=mode(...)
=Modalwert(...)

Median (Zentralwert): halbiert eine Stichprobe.

Anzahl der Daten der Stichprobe kleiner als Median =
= Anzahl der Daten der Stichprobe größer als Median

$$x_{\text{med}} = \begin{cases} x_{(n+1)/2} & \text{falls } n \text{ ungerade} \\ (x_{n/2} + x_{(n/2+1)})/2 & \text{falls } n \text{ gerade} \end{cases}$$

=median(...)
=Median(...)

Durchschnittswert (der arithmetische Mittelwert)

Beispiel: Schritte

$$x_1 + x_2 + x_3 =$$



$$= \bar{x} + \bar{x} + \bar{x} = 3\bar{x}$$

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} = 0$$

Die Summe der Abweichungen der Daten von diesem Wert ist gleich Null.

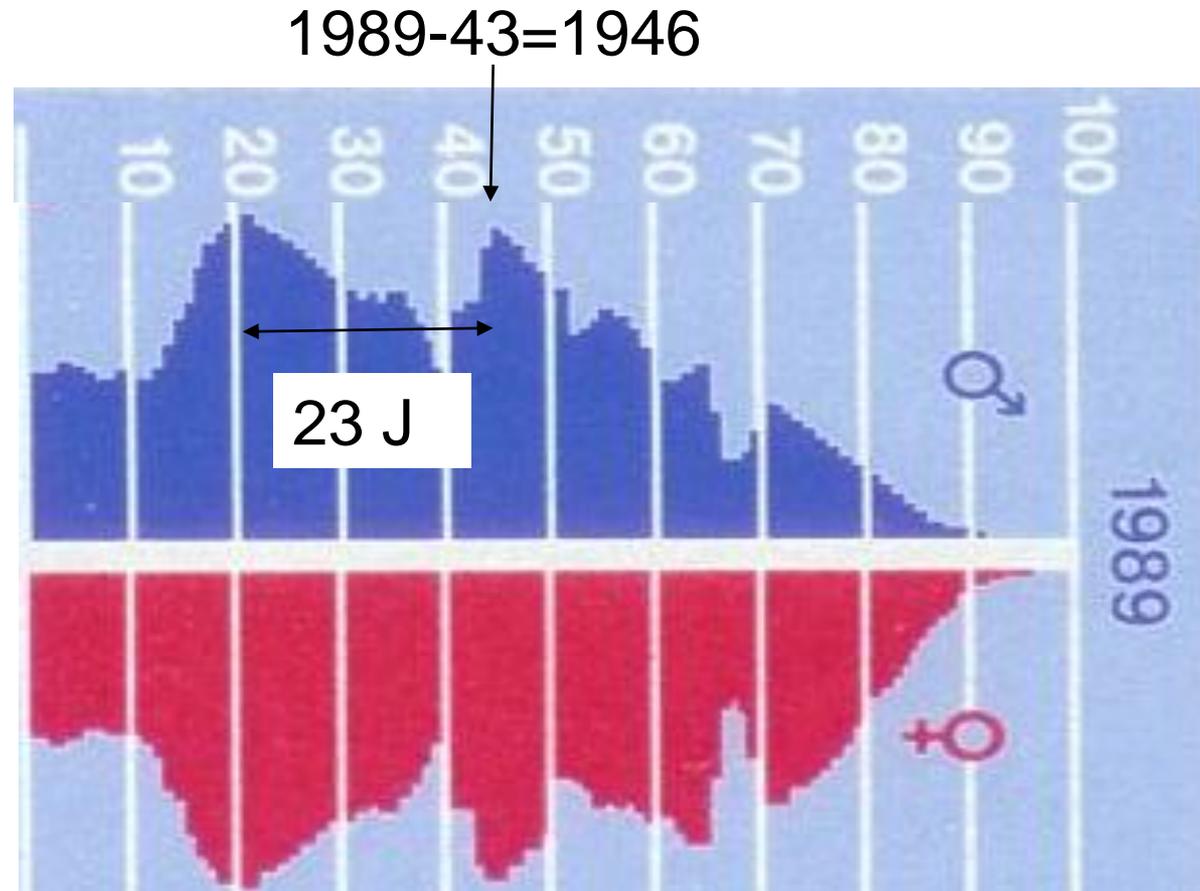
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

=Mittelwert(...)

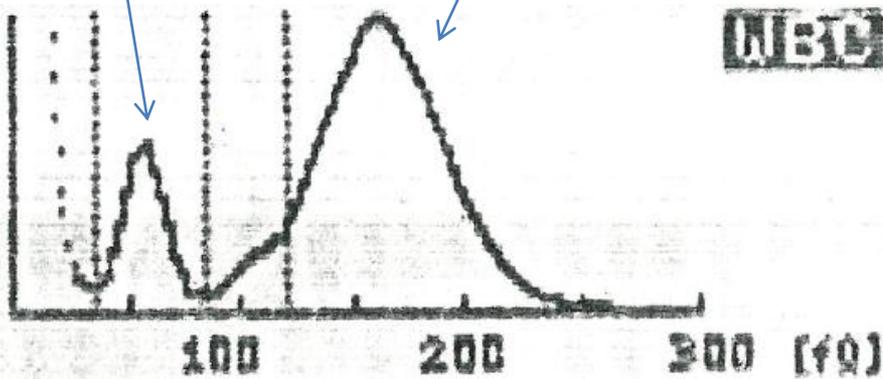
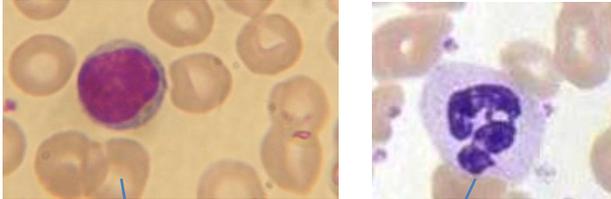
Beispiel, und modalität

Altersaufbau der deutschen Bevölkerung

- Unimodal:** die Verteilung hat nur einen Gipfel
- Bimodal:** die Verteilung hat zwei Gipfel.
- Multimodal:** die Verteilung hat mehrere Gipfel.



Beispiel in der Medizin

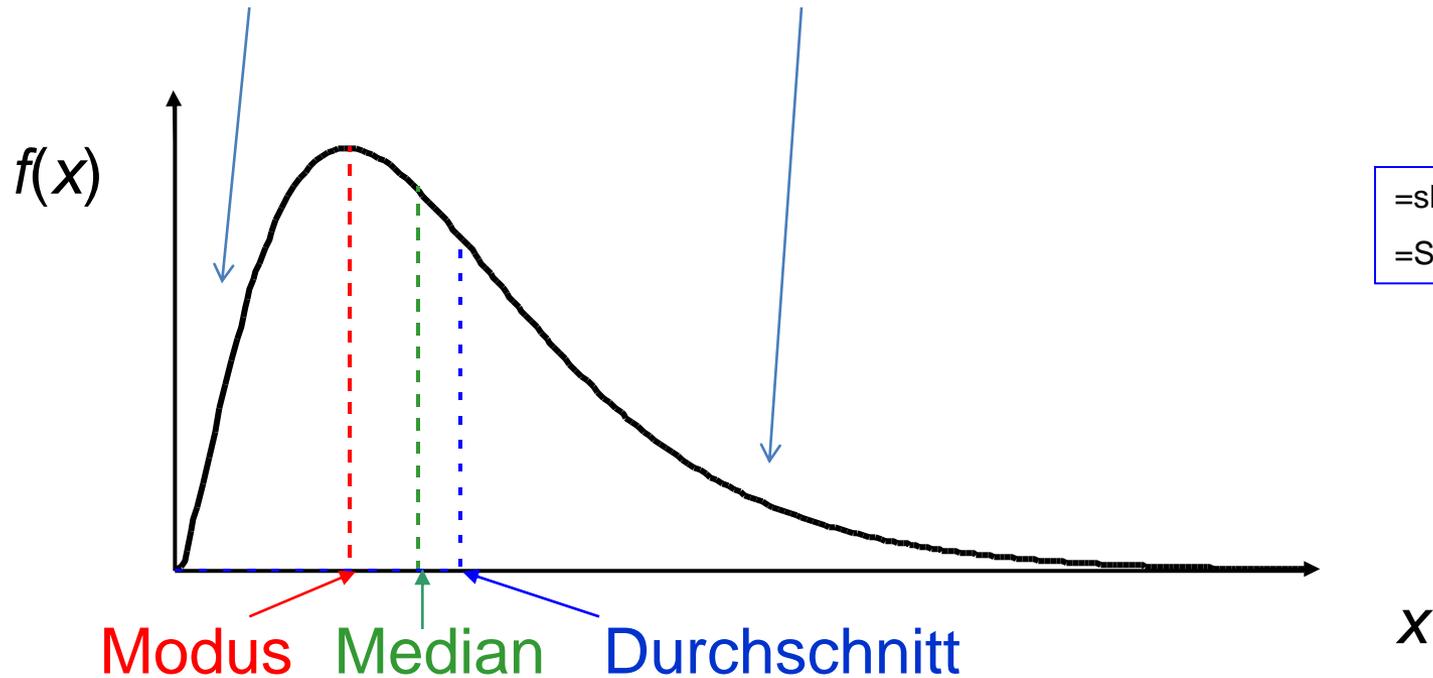


LYMPH%	16.2	%
MXD %	6.7	%
NEUT%	77.1	%
LYMPH#	1.2×10^3	/ μL
MXD #	0.5×10^3	/ μL
NEUT#	5.8×10^3	/ μL

Coulter Zähler

Größenverteilung der Zellen

Linkssteile bzw. **rechtsschiefe** Verteilung

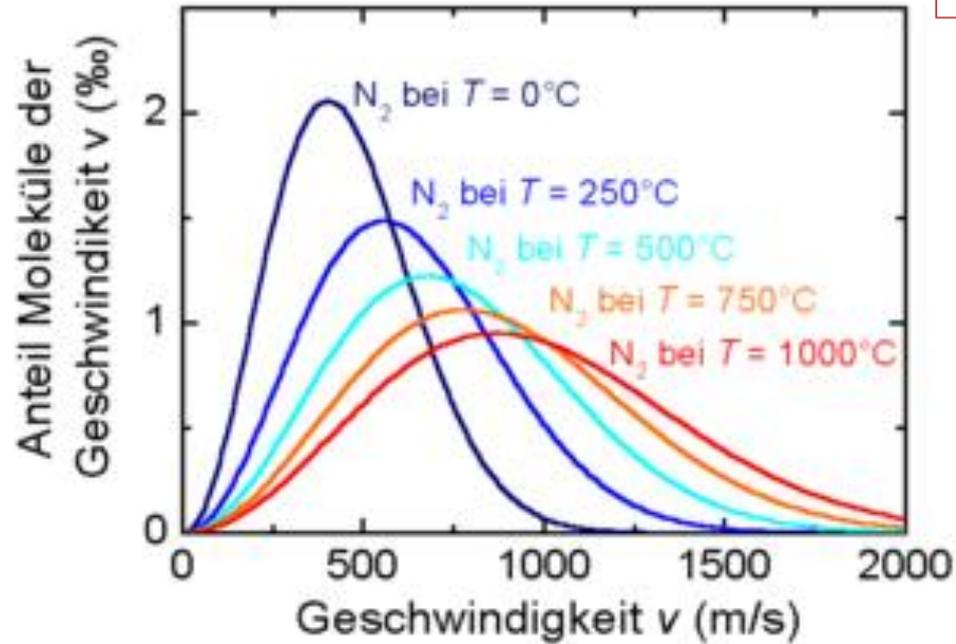


z.B. Einkommensverteilungen in einem Land:

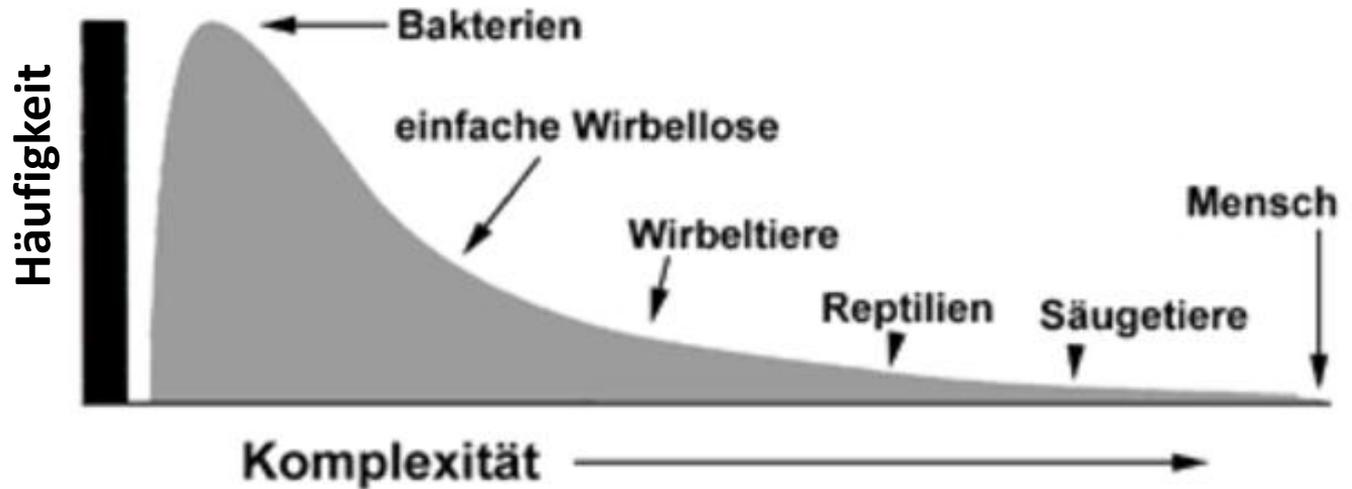
Der Großteil der Bevölkerung verdient relativ wenig, während es nur wenig Leute gibt, die sehr viel verdienen.

Weitere Beispiele

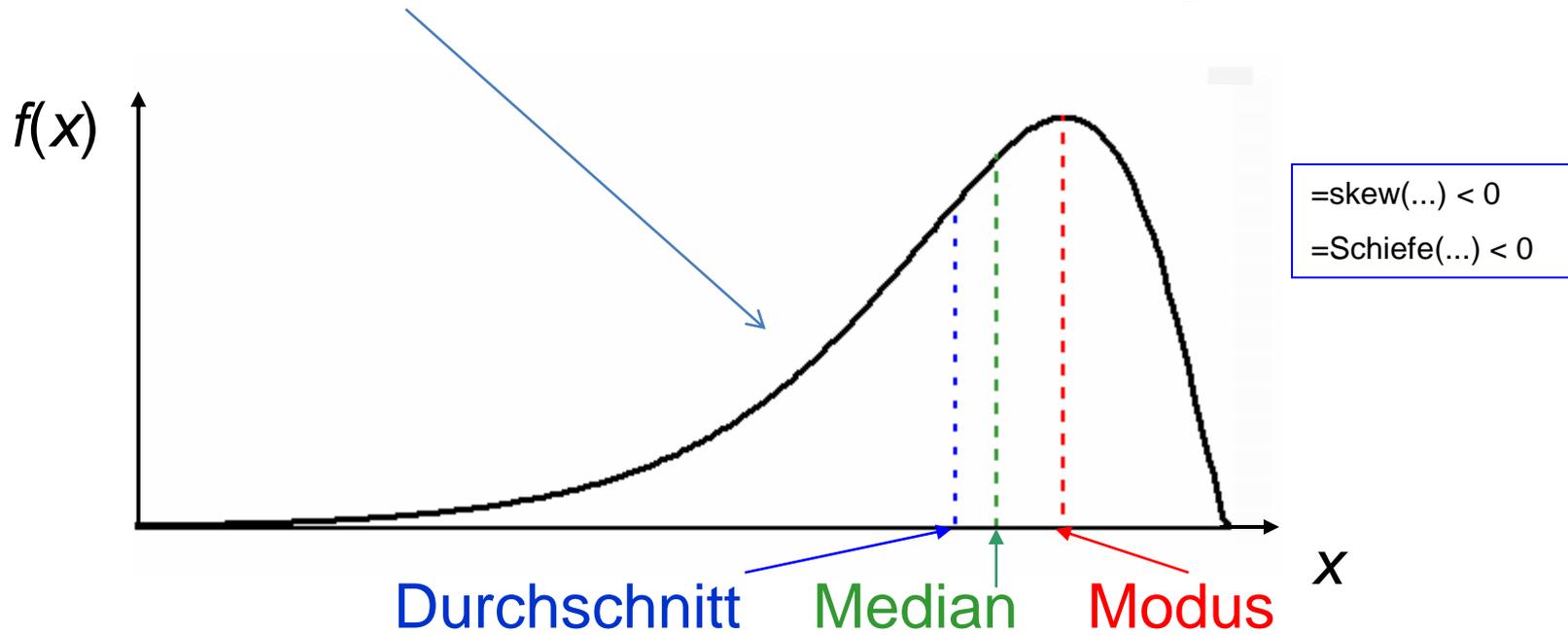
Maxwell-Boltzmann-Verteilur
(siehe später in der Physik)



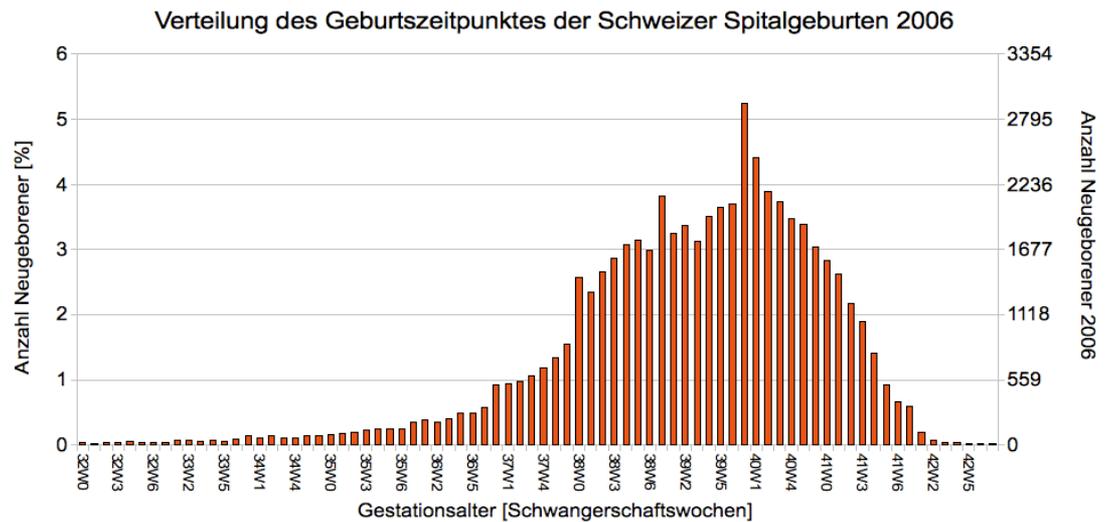
Komplexität der Tiere



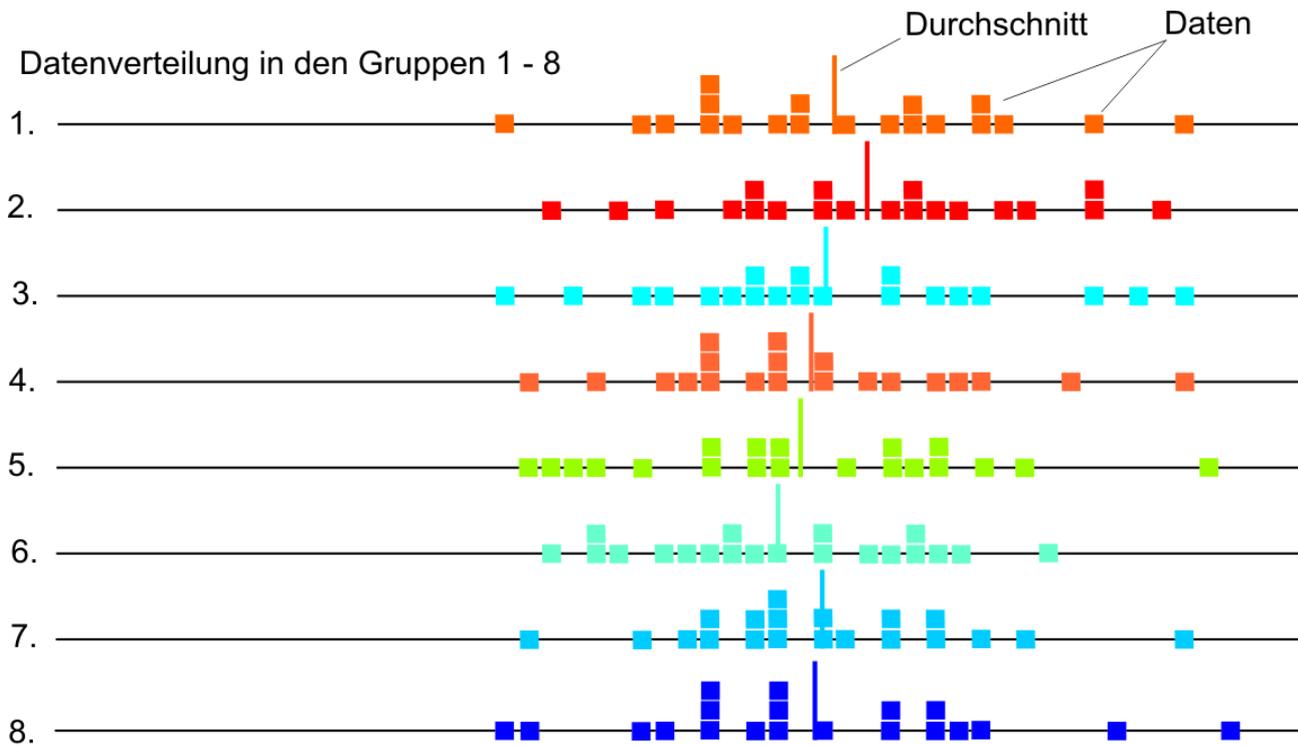
Linksschiefe bzw. rechtssteile Verteilung



z.B. Dauer einer Schwangerschaft



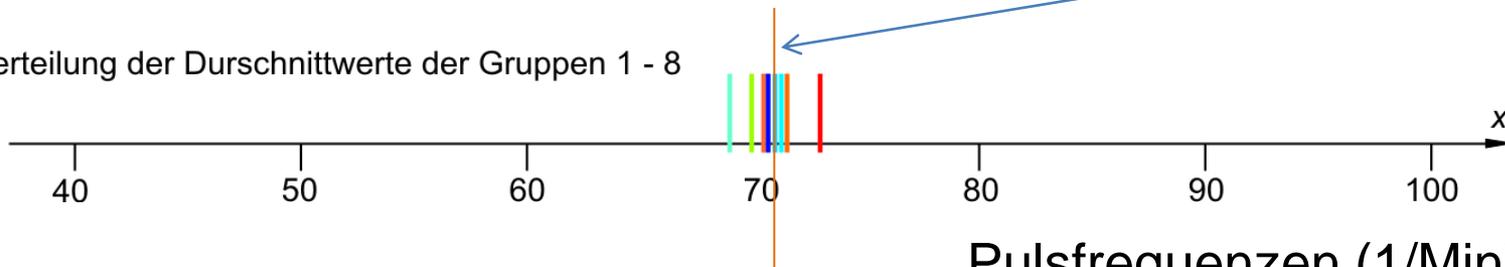
Daten und ihre Durchschnittswerte



Die Daten streuen um den Durchschnittswert.

Wenn wir mehrere Stichproben haben, dann **die Durchschnittswerte (Mittelwerte) streuen auch, und zwar um den Erwartungswert (μ)**

Verteilung der Durchschnittswerte der Gruppen 1 - 8



Streuungsparameter.

Charakterisierung der Variation der Daten

Standardabweichung

(Streuung der Messdaten, s):
die mittlere Abweichung vom
Durchschnitt:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

=stdev(...)
=Stabw(...)

das Quadrat der Streuung, die mittlere
quadratische Abweichung, auch als
Varianz bezeichnet:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

=var(...)
=Varianz(...)

Spannweite: $x_{\max} - x_{\min}$

=max(...)-min(...)

α -Quantil

$$0 < \alpha < 1$$

(seien dazu die x_i aufsteigend sortiert):

=Quantil(...)

Englisch: Percentile(...)

$$x_\alpha = \begin{cases} x_{[n\alpha]+1} & \text{falls } n\alpha \text{ keine ganze Zahl ist} \\ (x_{n\alpha} + x_{n\alpha+1})/2 & \text{falls } n\alpha \text{ ganzzahlig ist} \end{cases}$$

Beispiel: die x_i Werte sind ($i=1..12$), $n=12$:

0,4,4,6,7,9,10,11,13,15,17,20.

Wenn $\alpha=0.25$, dann $x_{0.25} = (4+6)/2=5$, weil $12*0.25=3$

Wenn $\alpha=0.3$, dann $x_{0.25} = 6$, weil $12*0.3=3.6$

mit Wörter: z.B. Quartile, Dezile, etc.

$x_{1/4}$ – unteres Quartil $x_{3/4}$ – oberes Quartil

$x_{1/10}$ – unteres Dezil $x_{9/10}$ – oberes Dezil

halber Quartilabstand : $(x_{3/4} - x_{1/4})/2$

=Quartile(...)

Hier kann nur
 $\alpha =$ einige
quartile sein!

Dezile

Durch Dezile (lat. „Zehntelwerte“) wird die Verteilung in 10 gleich große Teile zerlegt. Unterhalb des dritten Dezils liegen 30 % der Verteilung.

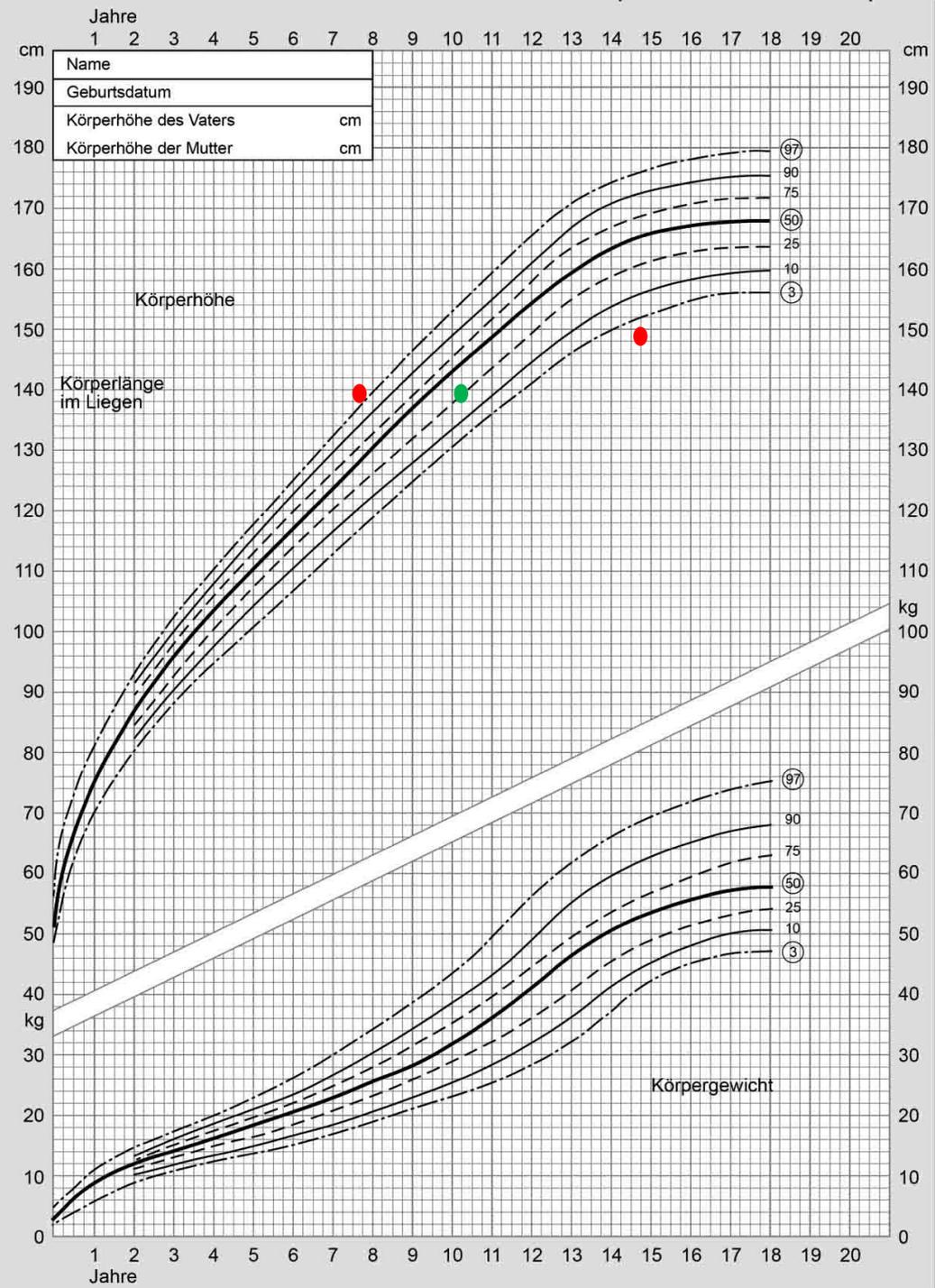
Perzentilenkurven sind ein Werkzeug für den Arzt.

Die **Fälle am Rahmen** sind einfach sichtbar.

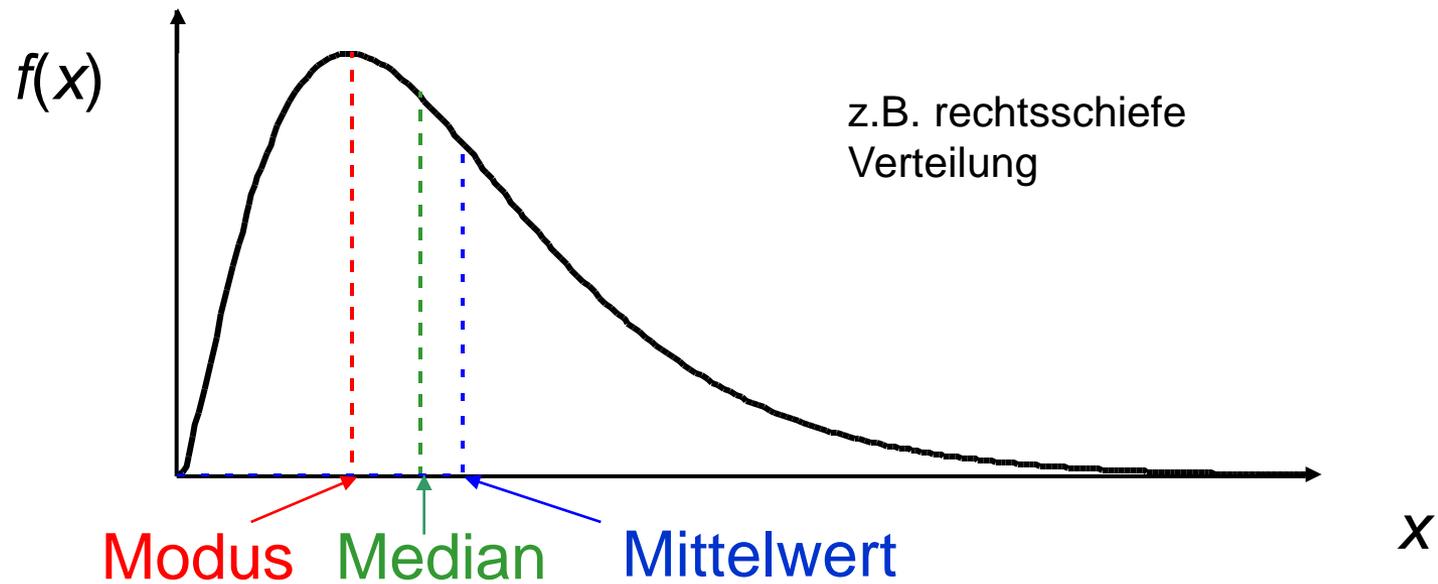
z.B.

Wachstums- und Gewichtskurven für Mädchen

=percentile(...)
=Quantil(...)



Die Lageparameter sind generell nicht identisch.
Bei einer symmetrischen Verteilung sind sie aber gleich.

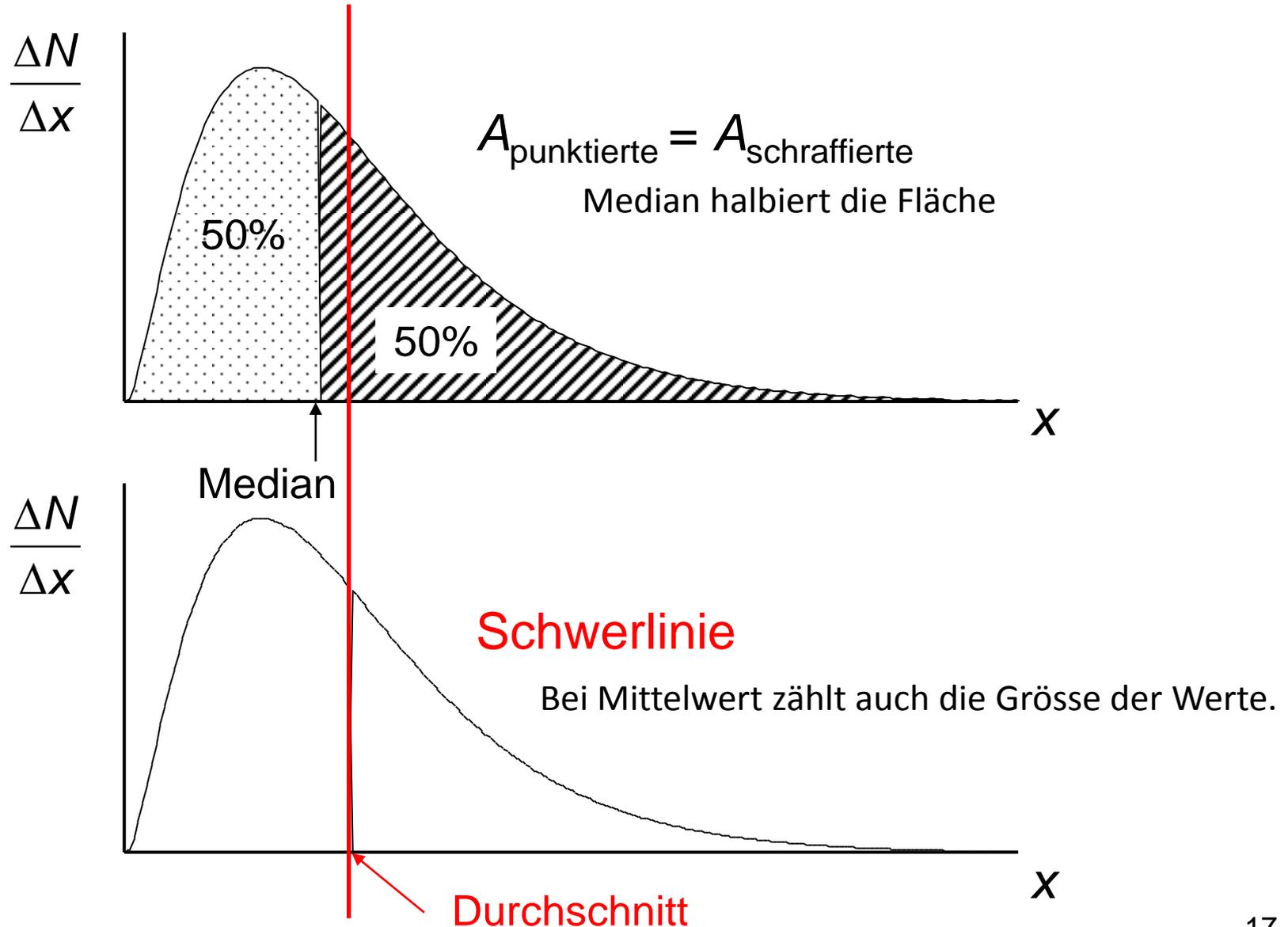


Vorsicht mit Skalentypen!

Besonders mit zahlen: die originelle Skala kann gut „nur“ nominal, oder ordinal sein (z.B. Noten)

Skalentypen	zulässige Lage-Parameter	zulässige Streuungs-Parameter
Nominalskala	Modus	–
Ordinalskala	Modus, Median	–
numerische Skalen	Modus, Median, Durchschnittswert	Spannweite, Quartilabstand, Standardabweichung

Lage des Medians und des Durchschnitts einer Verteilung

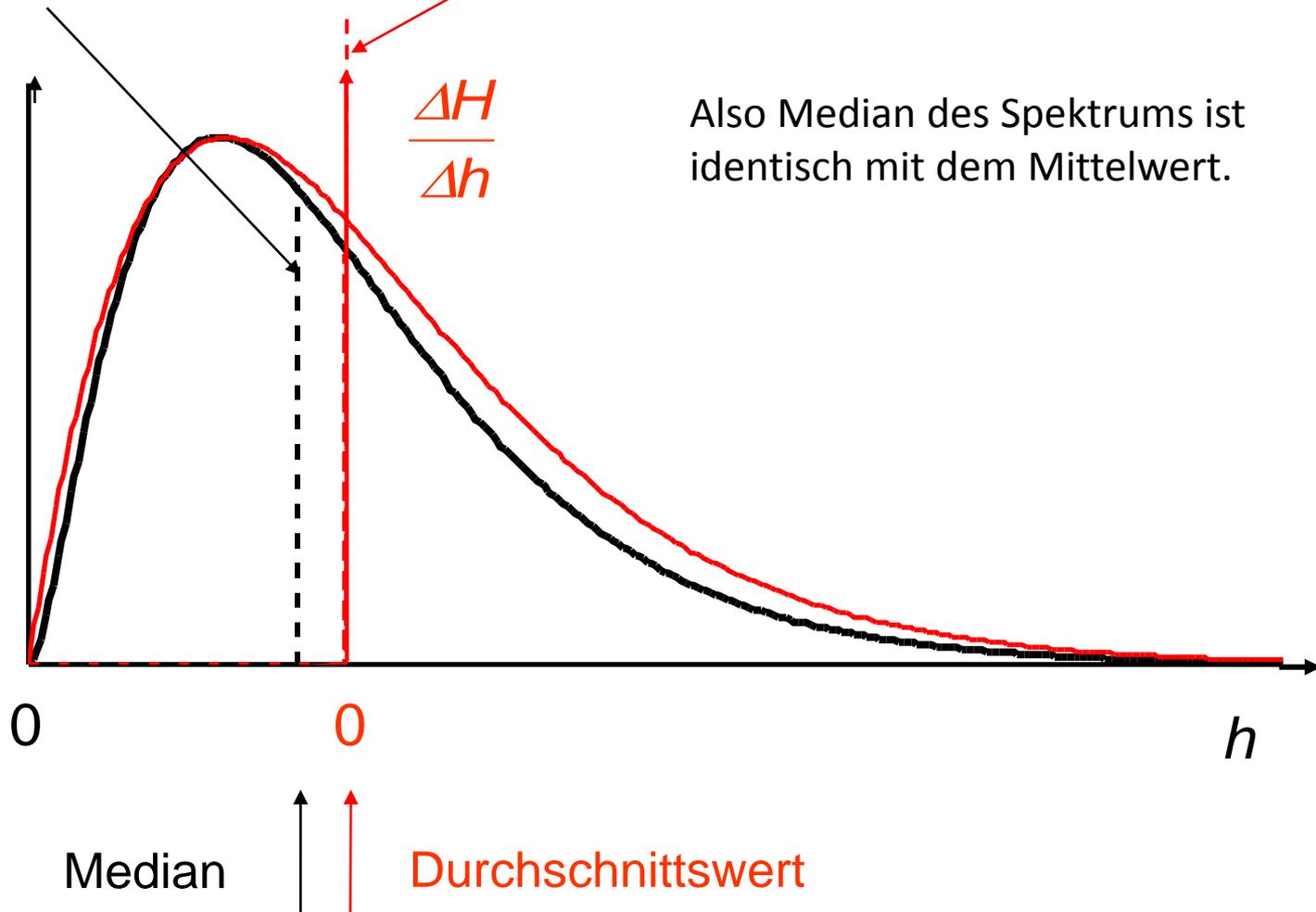


Mittelwert und Spektrumsmedian

Flächenhalbierungslinie der
HäufigkeitsverteilungFlächenhalbierungslinie des
Spektrums

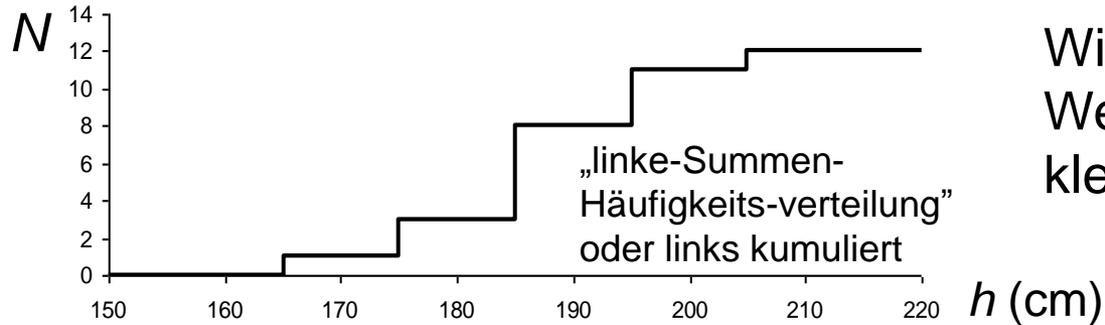
$$\frac{\Delta N}{\Delta h}$$

$$\frac{\Delta H}{\Delta h}$$

Also Median des Spektrums ist
identisch mit dem Mittelwert.

Summen- (kumulierte/kumulative) Häufigkeitsverteilung

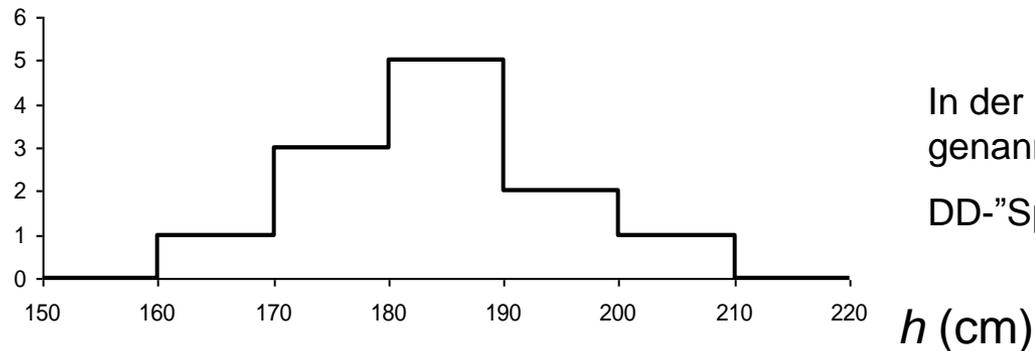
Summen-
Häufigkeits-
verteilung



Wieviele
Werte sind
kleiner als h ?

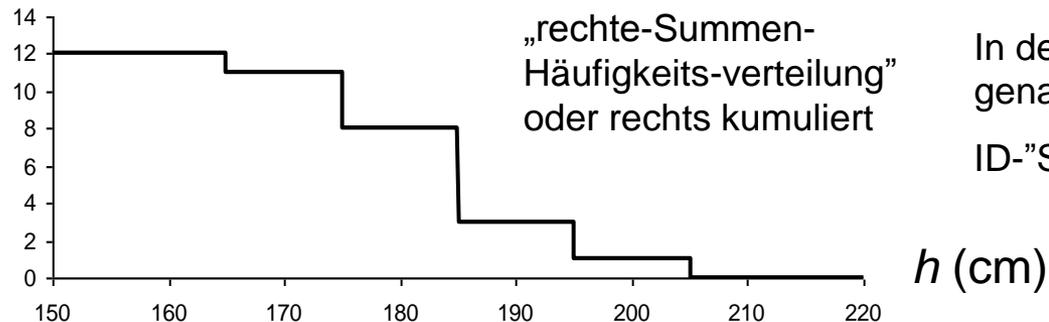
Häufigkeitsdichte-
Verteilung

$$\frac{\Delta N}{\Delta h} \left(\frac{1}{10\text{cm}} \right)$$



In der Messtechnik
genannt als
DD-„Spektrum“

$M = N_0 - N$
Wieviele
Werte sind
größer als h ?



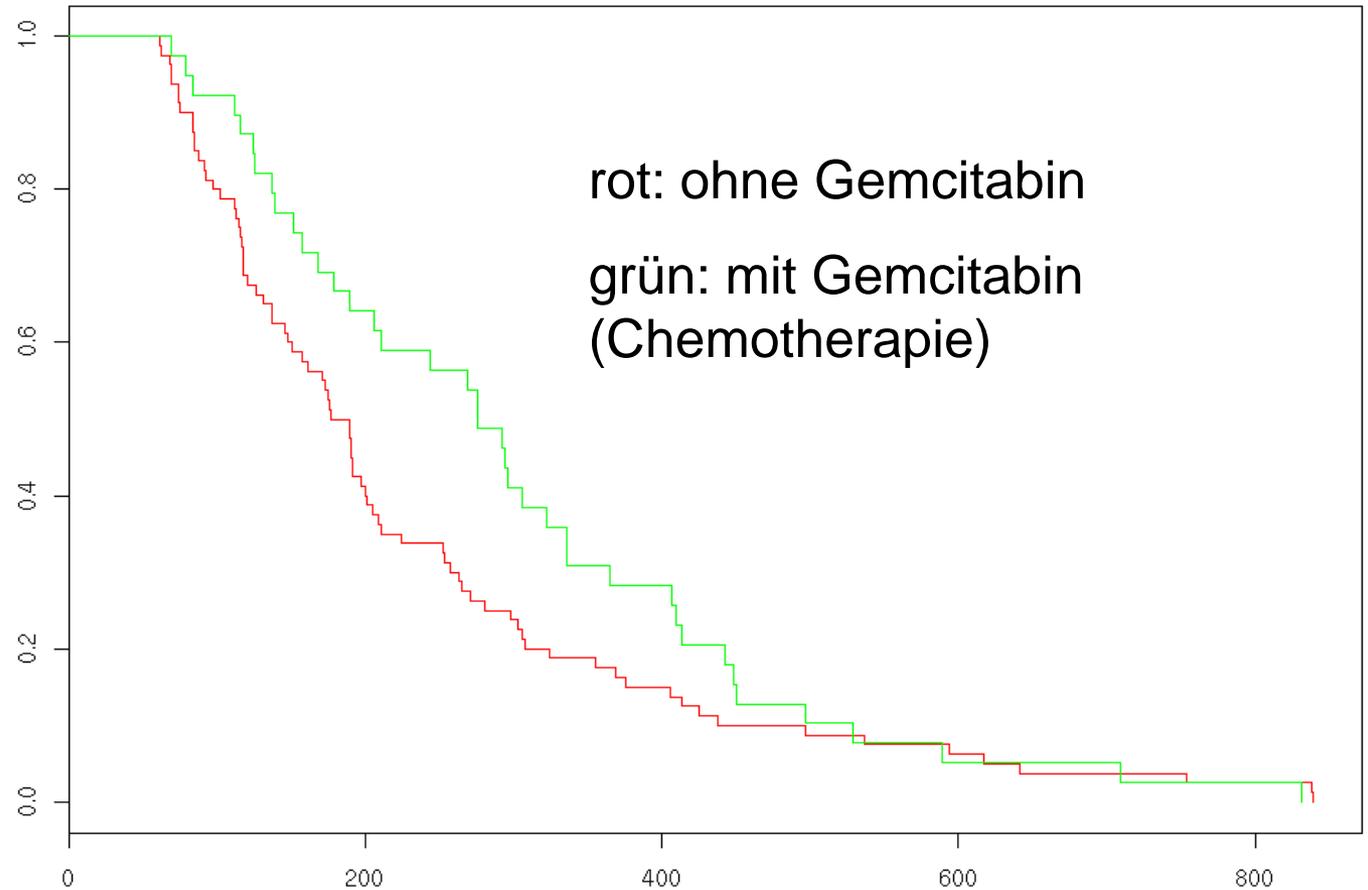
In der Messtechnik
genannt als
ID-„Spektrum“

Überlebenskurven

Wirkung der Chemotherapie. Pankreaskarzinom

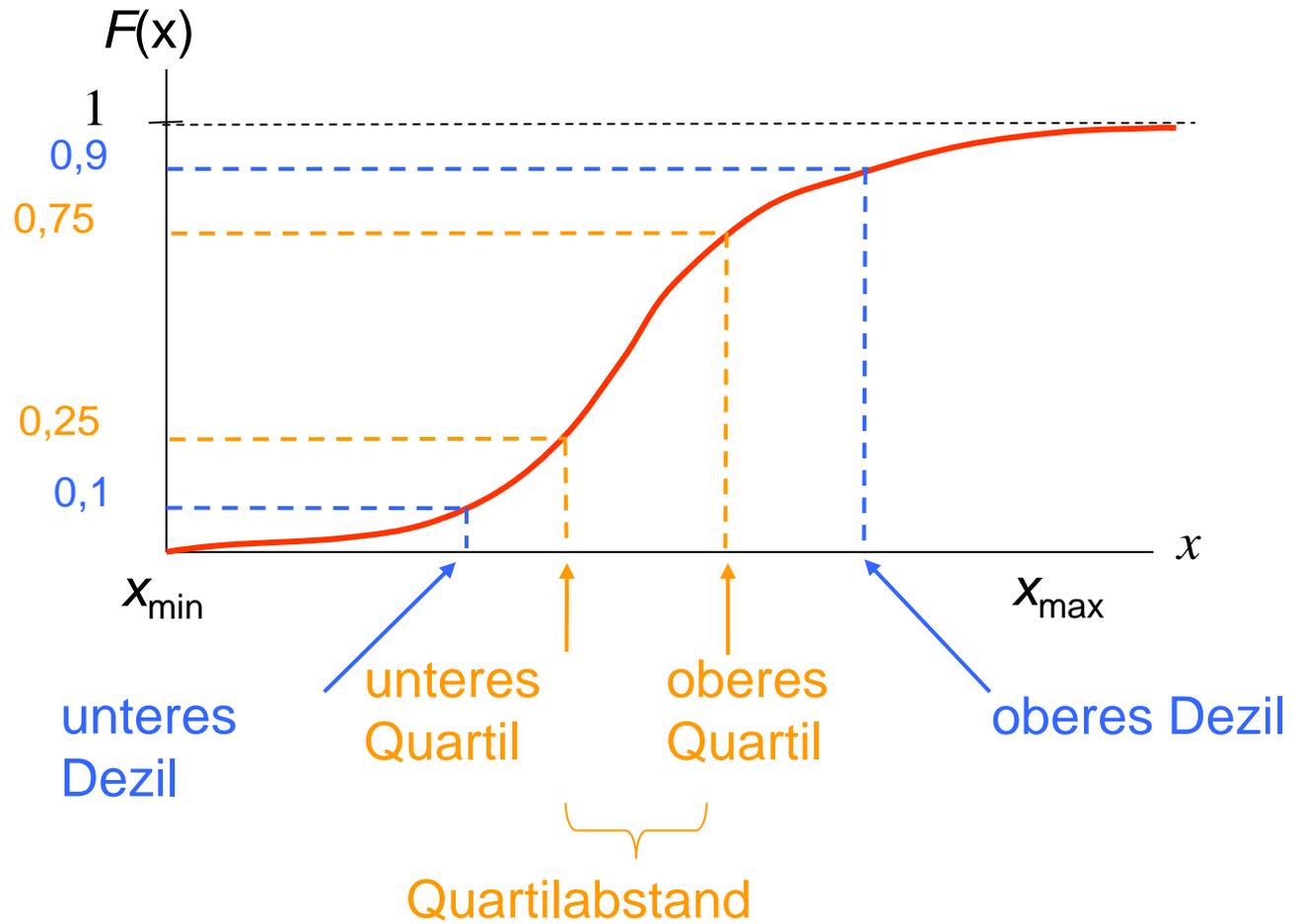
Relative rechts kumulierte Kurve

kumulatives Überleben nach der Operation

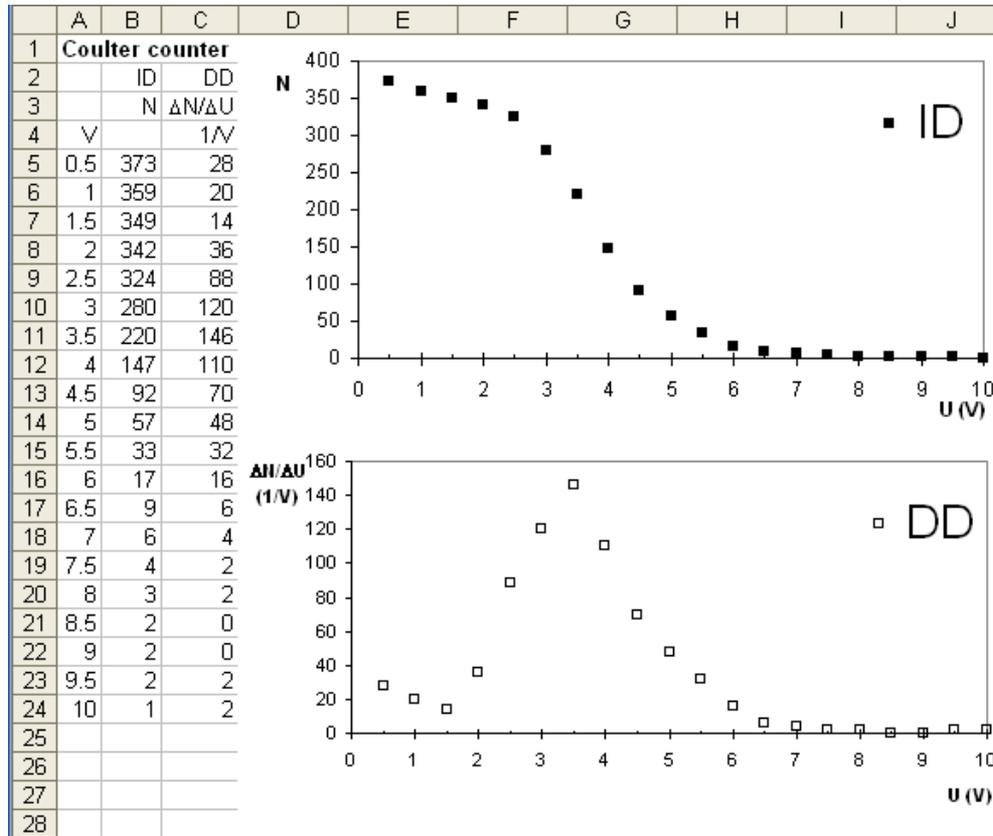


Überleben, Tage

Quantile und die relative Summenhäufigkeits-verteilung



**Beispiel in der
Physikpraktikum:
Coulter Zähler
(siehe viel später...)**



Verteilungen und Schätzungen

Grundlagen der Wahrscheinlichkeitslehre



Zufallsexperiment

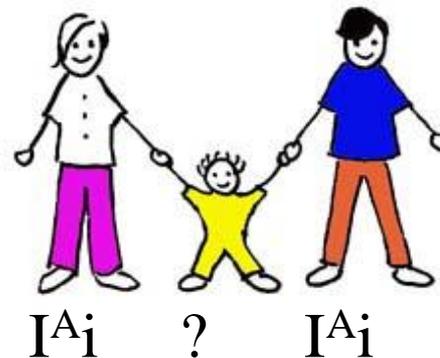
- Vorgang nach einer bestimmten Vorschrift ausgeführt
- (im Prinzip) beliebig oft wiederholbar
- sein Ergebnis ist zufallsabhängig (in der Natur ist es immer!)
Es gibt eine eingebaute Unsicherheit in der Natur.
- bei mehrmaligen Durchführung des Experiments beeinflussen die Ergebnisse einander nicht



Würfelspiel



Roulett



Blutgruppenversuch

Elementarereignisse

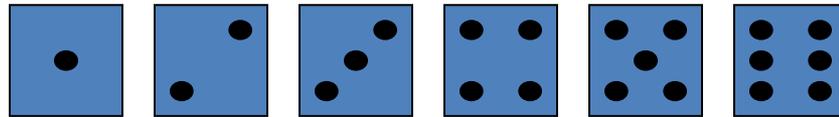
die einzelnen, nicht mehr zerlegbaren und sich gegenseitig ausschliessenden Ausgänge oder Ergebnisse eines Zufallsexperimentes

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“: $\Omega = \{I^A I^A, I^A i, i I^A, ii\}$

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:

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

Dabei denken wir, dass 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

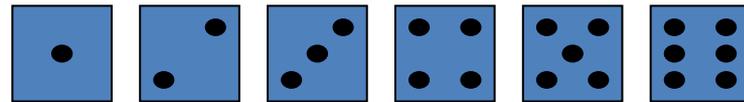
$$p(E) = \frac{g}{m}$$

← günstig

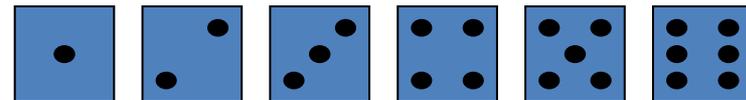
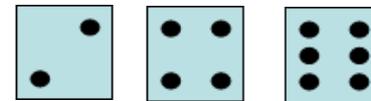
← alle

Würfelexperiment:

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



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



Münzenexperiment:

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



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, $p(A)$:

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

Wenn $n \rightarrow$ unendlich

Eigenschaften der Wahrscheinlichkeit

→ $0 \leq p(A) \leq 1$

→ $p(\text{sicheres Ereignis}) = 1$

→ $p(\text{unmögliches Ereignis}) = 0$

Verteilungen

Population



Wahrscheinlichkeitsverteilung

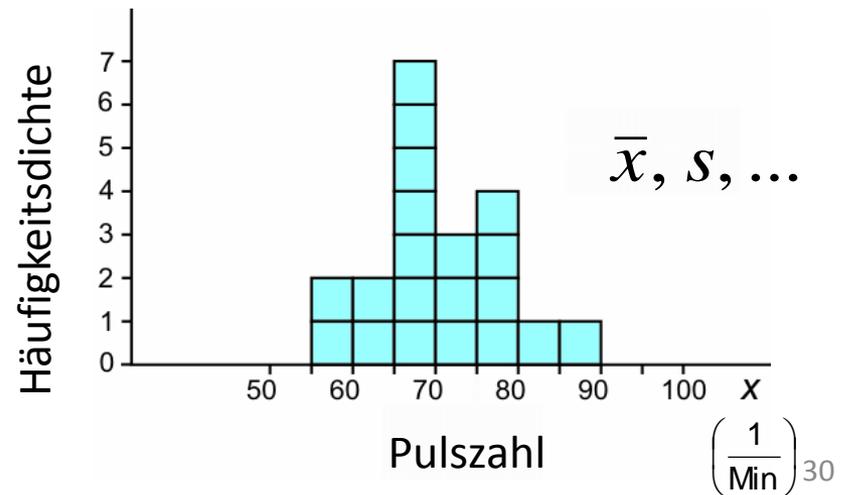


?, ?, ...

Stichprobe



$$\frac{\Delta n}{\Delta x} \left(\frac{\text{Min}}{5} \right)$$

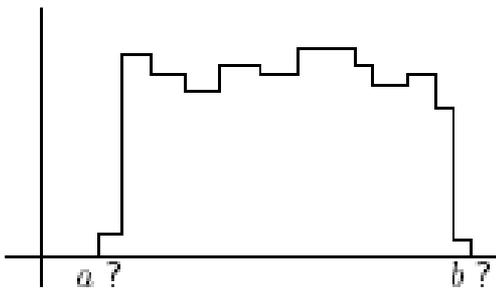


Verteilungen

Wie kann man die theoretische Verteilung bestimmen?

Vermutung

(nach dem Histogramm)



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
- ...

- **kontinuierliche Verteilungen**

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

diskrete Zufallsgröße

zB: Anzahl der Kranken,
Augenzahl des Würfels

kontinuierliche Zufallsgröße

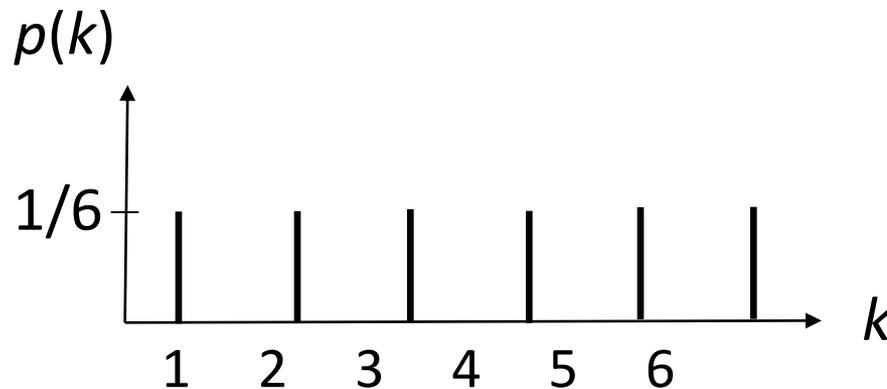
zB: Blutdruck, Körperhöhe,...

Diskrete Gleichverteilung



Beispiel:

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



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

weitere Beispiele:

Münzenversuch



Würfelexperiment
mit einem Ikosaeder



Lageparameter der Verteilung

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

$$\mu = \sum_i x_i p(x_i)$$

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

$$\mu = \sum_i x_i p(x_i)$$

$$\bar{x} = \sum_i x_i h_i$$

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

Rel.Häufigkeit

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

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

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

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

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

x_i : Augenzahl

n_i : absolute Häufigkeit

h_i : relative Häufigkeit

$$\bar{x} \xrightarrow{n \rightarrow \infty} \mu$$

Streuung der Verteilung

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

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

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

$$S \xrightarrow{n \rightarrow \infty} \sigma$$

empirische → theoretische
Streuung Streuung

(Standardabweichung)

Normalverteilung

Verteilungsdichtefunktion:

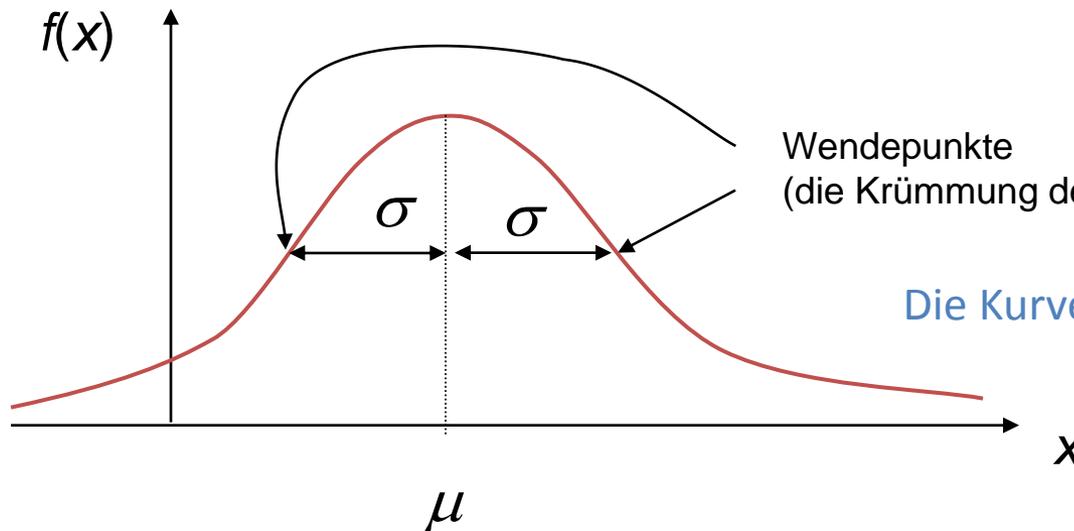
Parameter der Normalverteilung:

Erwartungswert: μ

Streuung: σ

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

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



Die Kurve geht aber NIE ganz zu Null!

Normalverteilung (Gauss-Verteilung)

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



DL0998939U1



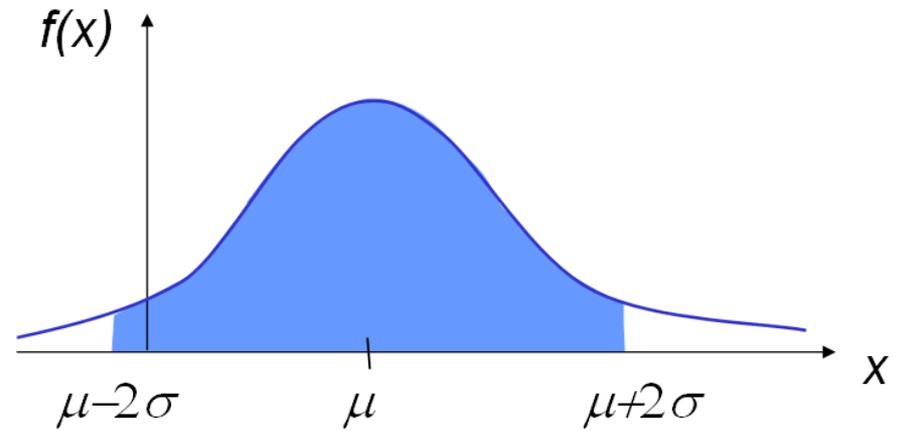
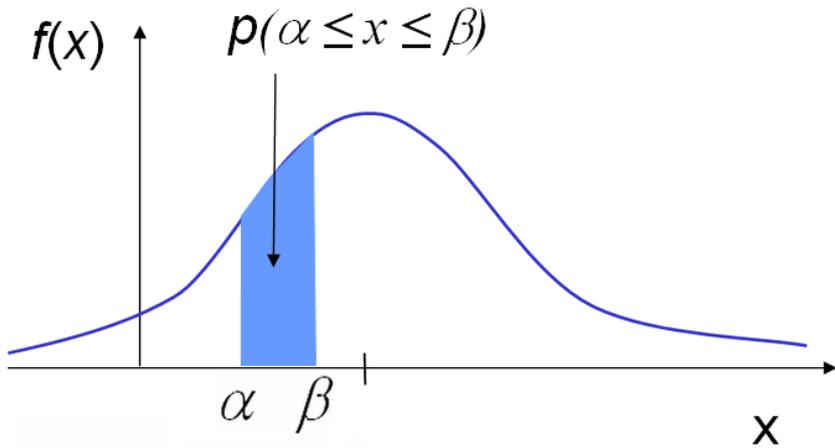
Deutsche Bundesbank

Kirsten Jansen
Frankfurt am Main
1. Oktober 1993

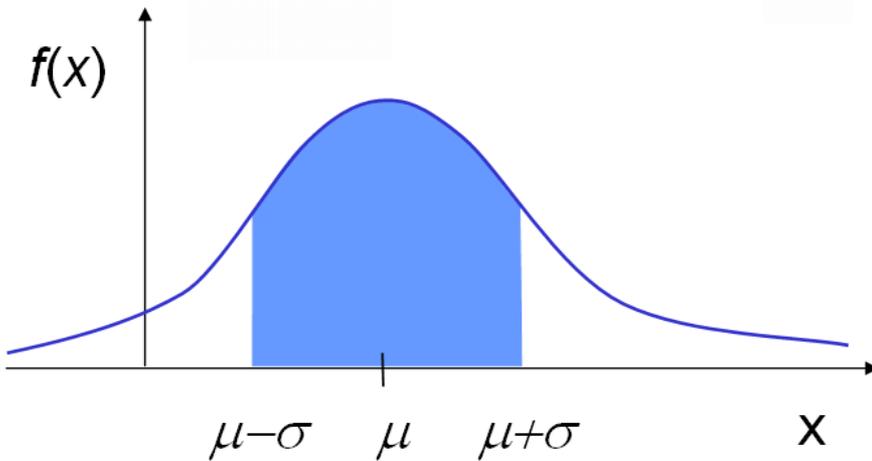


Normalverteilung

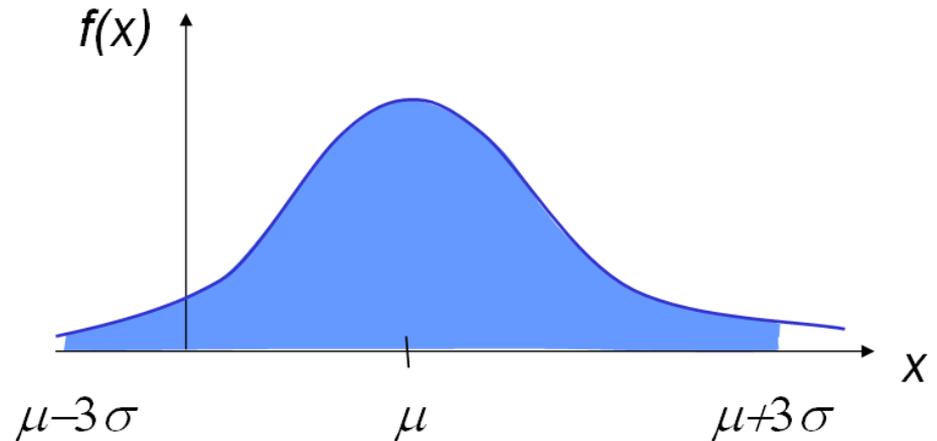
Wahrscheinlichkeit ist eine Oberfläche unter der Dichtefunktion!



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



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



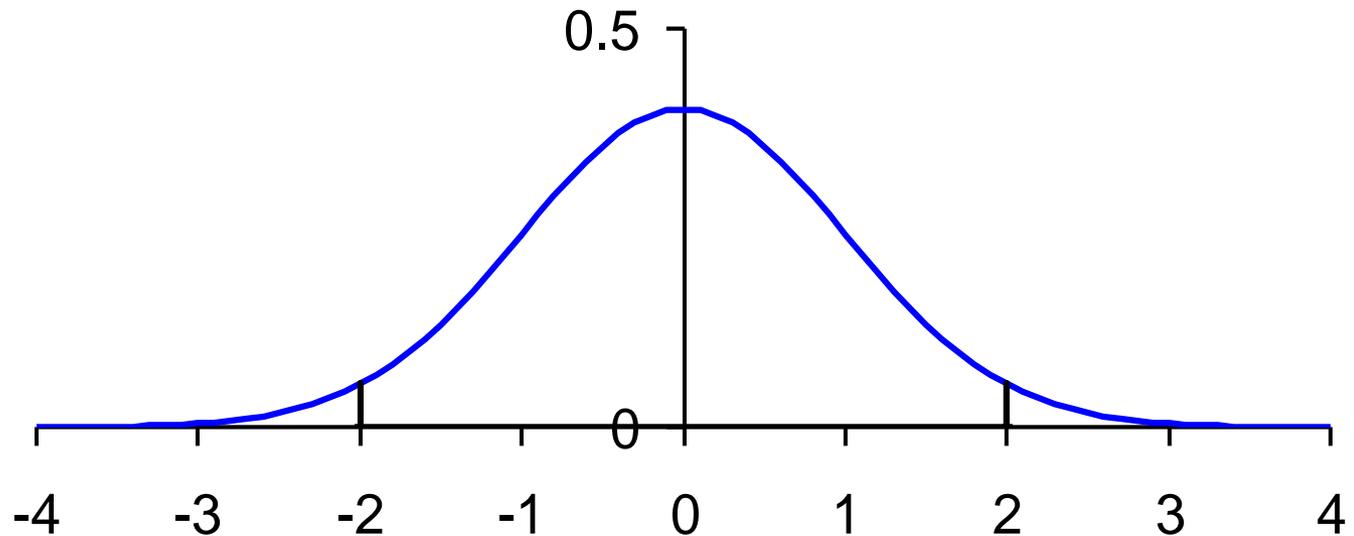
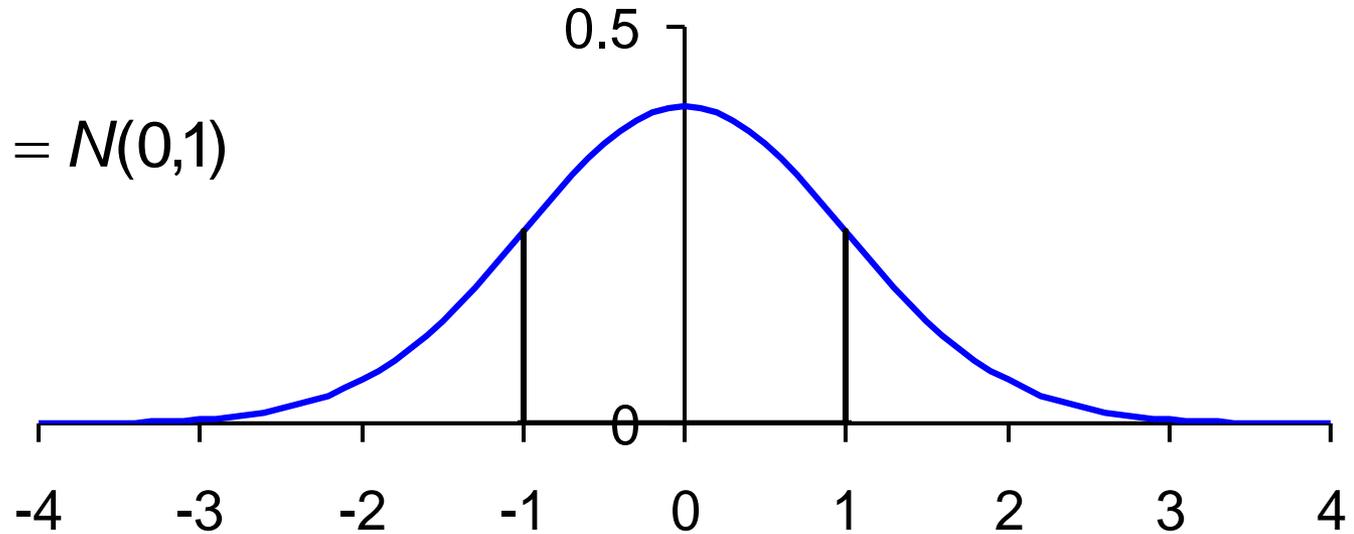
$$p(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 99,8\%$$

Standard - Normalverteilung

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

$$\mu = 0$$

$$\sigma = 1$$



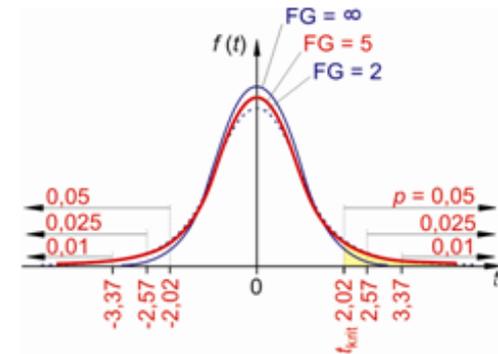
1. STATISTISCHE TABELLEN

t-VERTEILUNG

Freiheitsgrad (FG)	p (Irrtumswahrscheinlichkeit, einseitiger Test)						
	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

t-Verteilungsfamilie



„Glockenkurven“

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

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

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



Zentraler Grenzwertsatz

- Es seien x_1, x_2, \dots, x_n unabhängige Zufallsgrößen, die alle derselben Verteilung haben.

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

- 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

$N = \text{„unendlich“}$

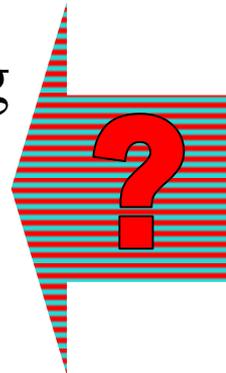


Stichprobe

$n = \text{endlich}$



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!

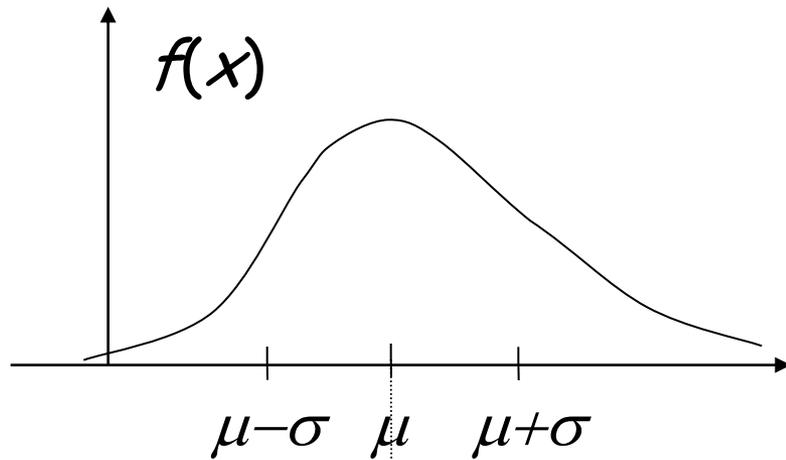
Intervallschätzungen

- 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.

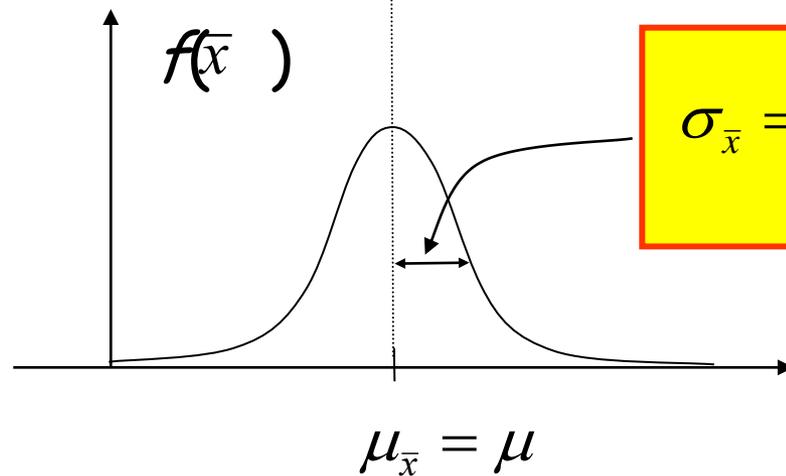


Zb.: Erwartungswert der Pulszahl ist bei
95% Konfidenzniveau: 74 ± 6 ^{1/Min}

Konfidenzintervall für den Erwartungswert



x zB: Körperhöhe

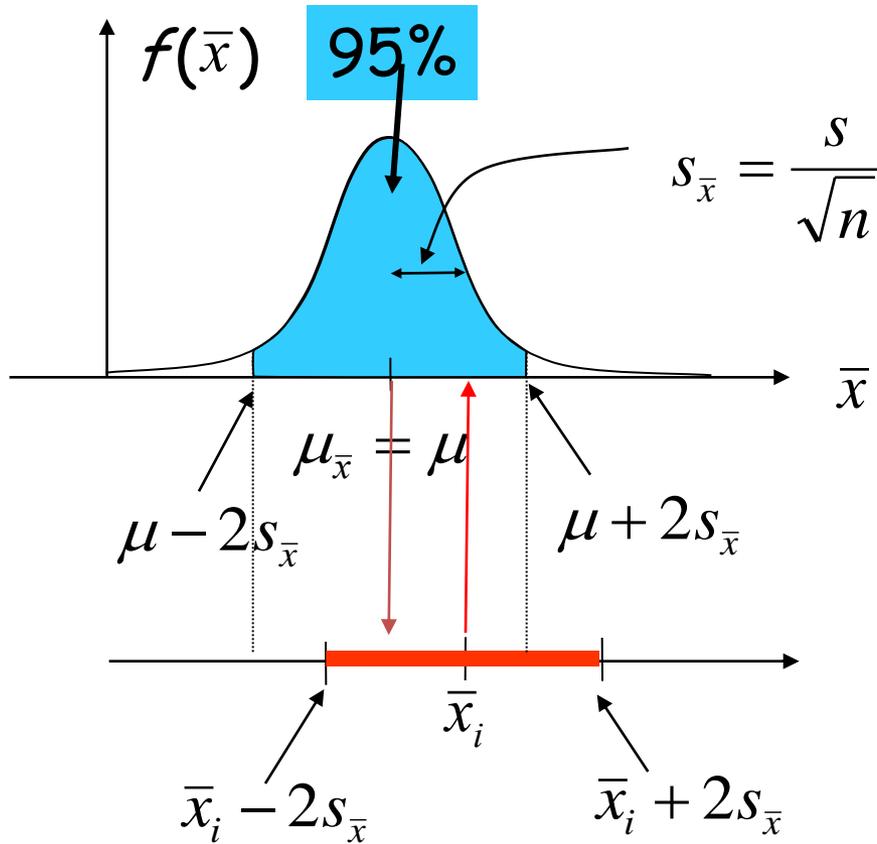


$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx s_{\bar{x}}$$

Standardfehler

\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

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

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

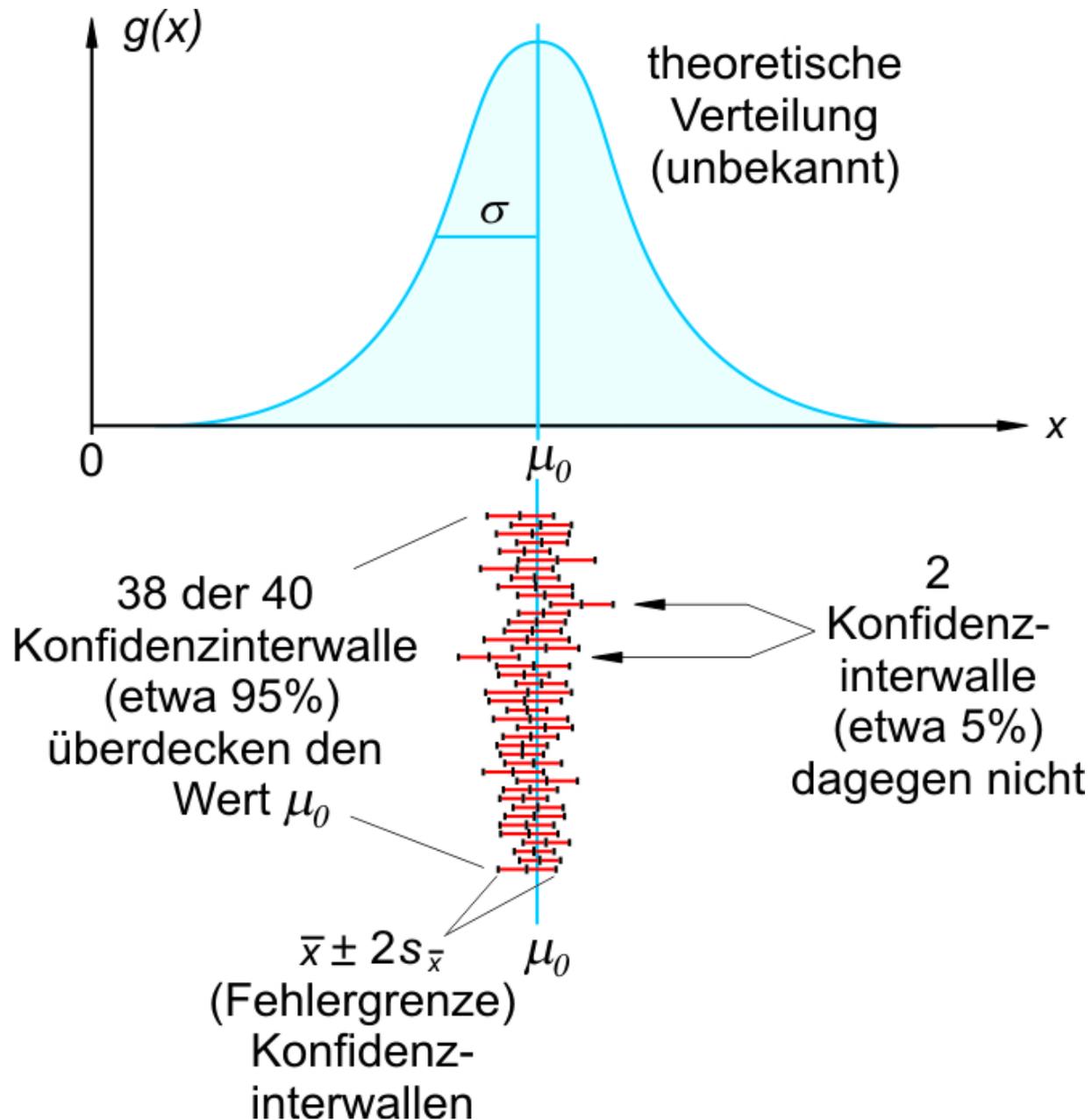
- mit 68% Wahrscheinlichkeit im Intervall: $\bar{x} - s_{\bar{x}}, \bar{x} + s_{\bar{x}}$

- 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 $s_{\bar{x}} \rightarrow 0$



Zusammenfassung der Schätzungen

- Punktsätzungen:

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

Intervallschätzung
für den Erwartungswert:

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