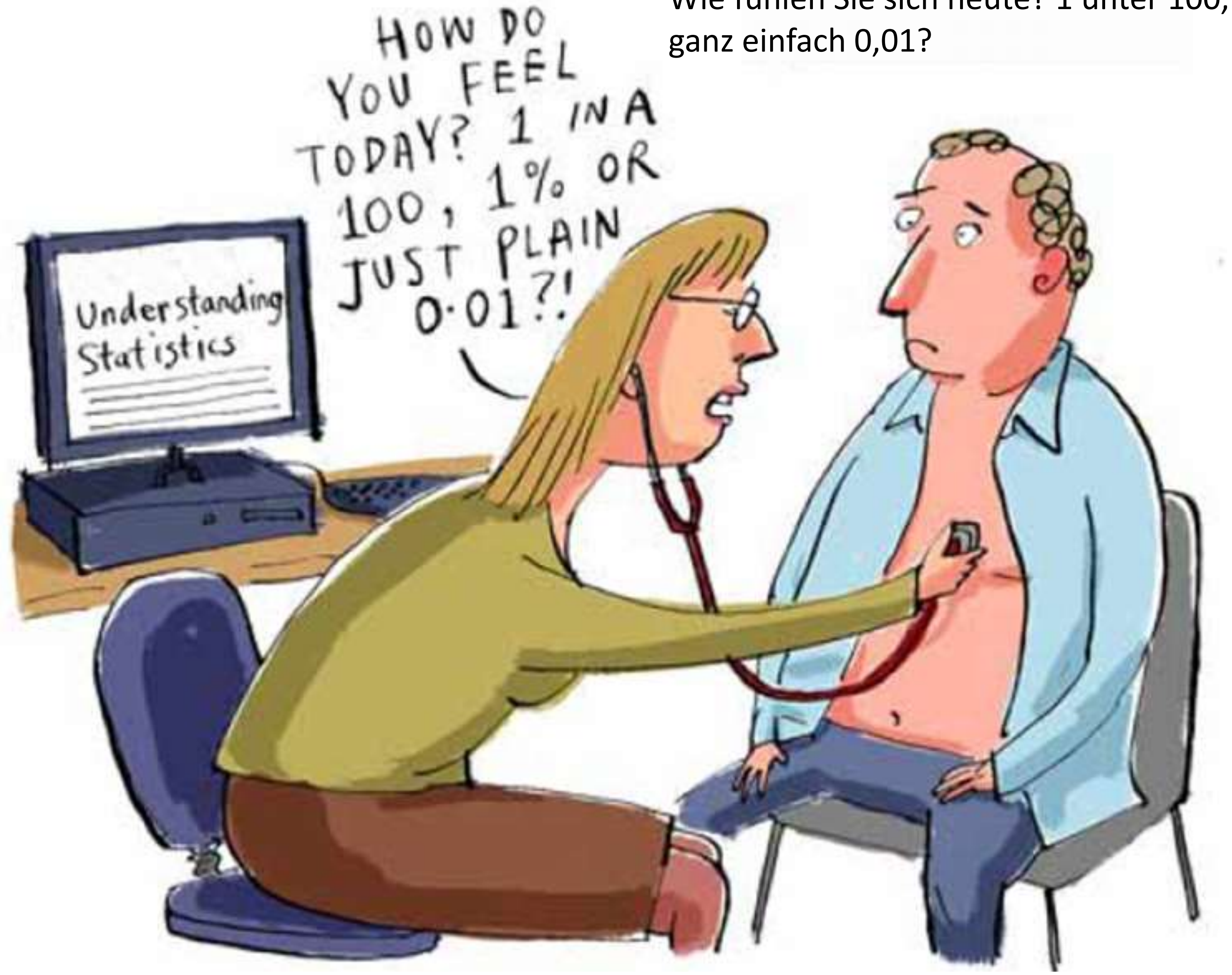
Wie fühlen Sie sich heute? 1 unter 100, 1%, oder ganz einfach 0,01?



Deskriptive Statistik

Schay G.

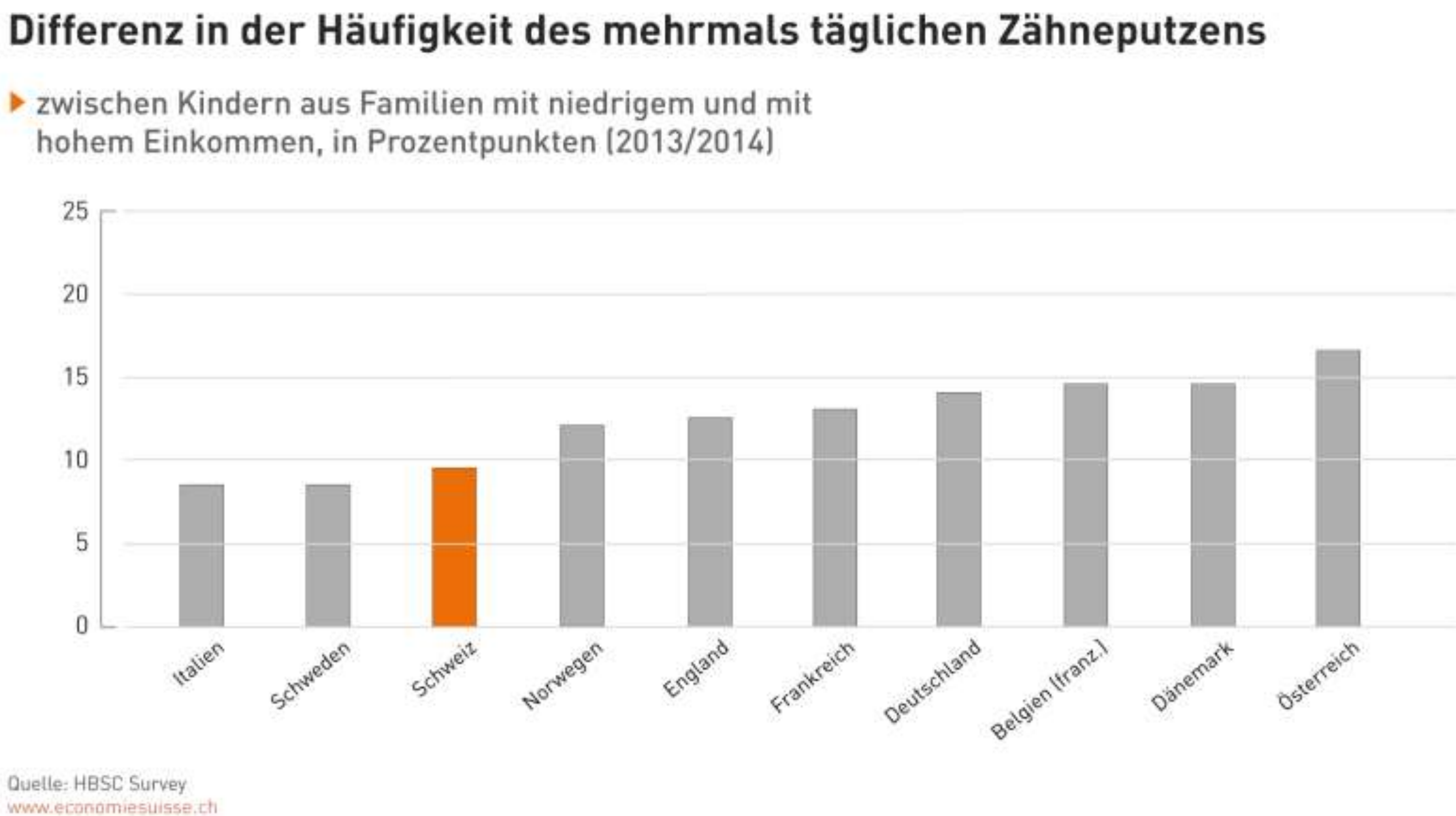
Wie können Sie effizient naturwissenchaften lernen?

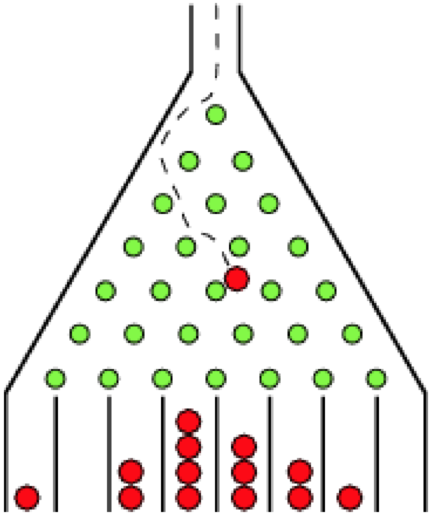
Naturwissenschaft ist wie eine Sprache: am besten übt man, und versucht neue dinge zu verstehen.
Was unverständlich scheint, das schaut man nochmal durch, solange bis man es wenigstens ein bischen benutzen kann.

Lassen Sie nie eine Frage offen länger als eine Woche!
Sie können während der Vorlesung und Praktika ihre Fragen stellen, wir versuchen die zu beantworten.

Kommen sie zu der Vorlesungen mit
unter
http://biofiz.semmelweis.hu erreich
Falls nicht in dem aktuellen Semeste
vorigen Jahr mit, die Veränderungen

Statistik ist auch in der Zahnmedizin…





Die Statistik beschäftigt sich mit
Massenerscheinungen,
aber

Einzelereignisse sind am meisten zufällig

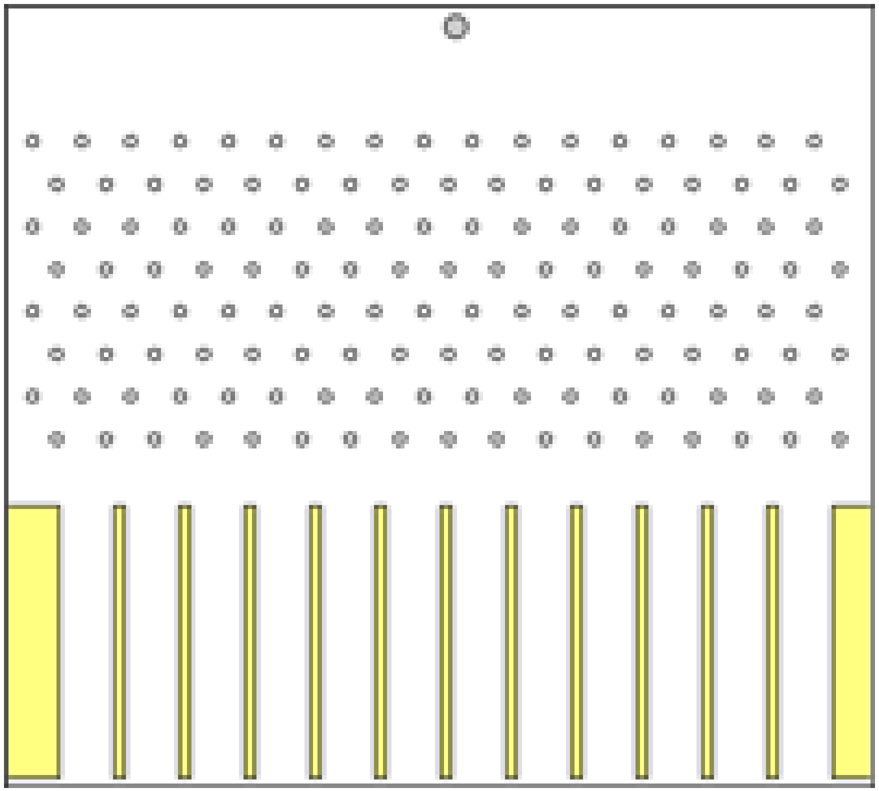
Statistik benutzt die Methoden der
Wahrscheinlichkeitsrechnung.

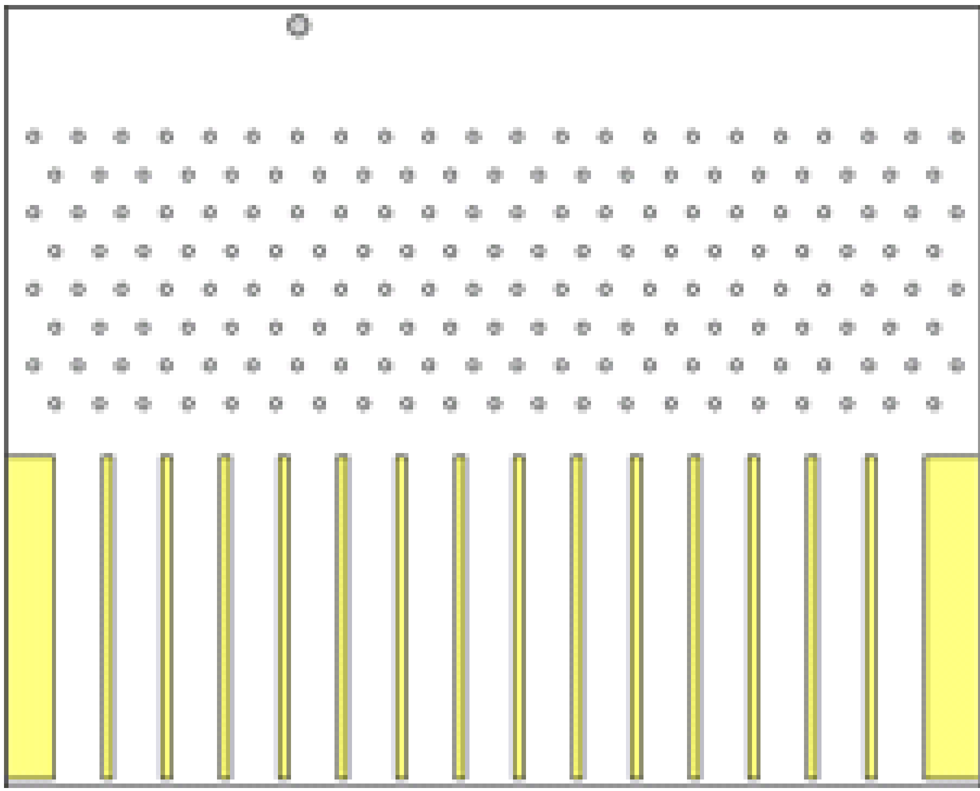
Fundamentalregeln:

Statistischen Aussagen beziehen sich nie auf ein Einzelereignis, sondern nur **auf Gesamtheiten vieler Ereignisse.**

Jede statistische Aussage ist mit einer **prinzipiell unvermeidlichen Unsicherheit** behaftet.

Galtonscher Brett



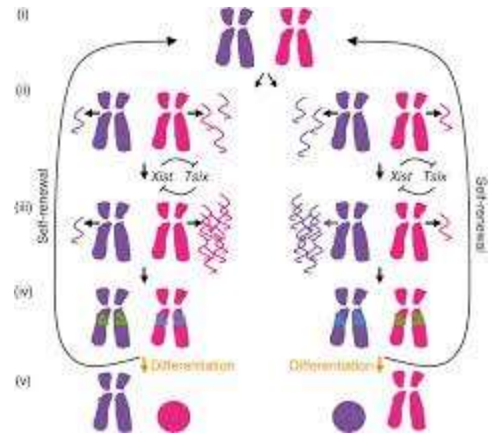


Die einzelnen Kugeln können wir theoretisch folgen, aber doch nicht vorhersagen welchen \
trotzdem, ein durchschnittliches Benehmen können wir beobachten.

http://mw.concord.org/modeler/showcase/mechanics/galton1.gif
http://mw.concord.org/modeler/showcase/mechanics/galton4.gif

Zufälligkeit ist überall in der Natur

Federteile und Blütenblätter



Gene

Blitze

Doch: wir sehen auch Ordnung...

Blätter

Statistik versucht Konzepte hinter, und in der Zufälligkeiten zu finden.

Oft „das grosse Bild” zeigt etwas „sinnvolles”, wobei die einzelnen Elemente anscheinend rein zufällig sind.

Wozu braucht eine Ärztin / ein Arzt Statistik?

— zum Verstehen der medizinischen Fachliteratur („How to Read a Paper“) insbesondere von Originalarbeiten in Fachzeitschriften über
  - experimentelle
  - klinische
  - epidemiologische
  - sonstige (z. B. gesundheitsökonomische) Studien

— „Evidence-based Medicine“ Bewertung und Kommunikation von Chancen und Risiken

— bei eigenen Untersuchungen
  - Doktorarbeit
  - Industrie
  - Gesundheitsbehörden

http://www.medizin.uni-koeln.de/kai/imsie/kursinfo/q1/Q1-01-Einfuehrung.pdf

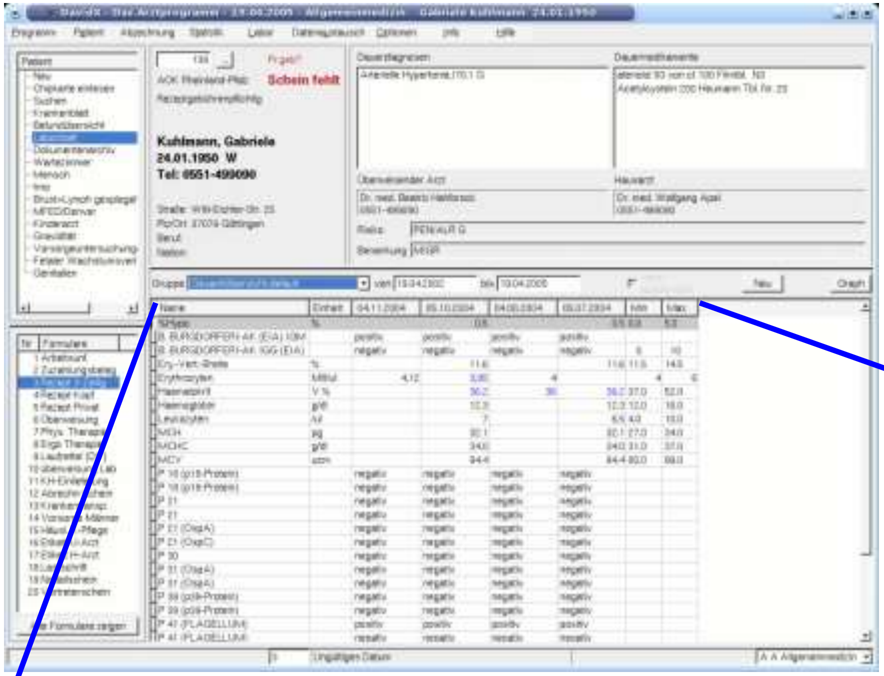
das erste Anwendungs-
gebiet der Statistik
bestand in der
Staatsbeschreibung
(Völkszählung)
Status = Zustand

Semmelweis (1818-
1865) war der erste
bekannte Arzt, der
den Nutzen einer
neuen Therapie
mit statistischen
Methoden belegte

Was messen Physiker, Arzt und Medizinstudent?

WER MISST WAS?		
PHYSIKER	ARZT	MEDIZINSTUDENT IM PHYSIKPRAKTIKUM
Länge	Körpergröße	Durchmesser von Erythrozyten (3)
Frequenz	Pulsfrequens	Impulshäufigkeit (9,20)
Temperatur	Körpertemperatur	—
Konzentration	Blutzuckerspiegel	Glycerinkoncentration der Lösung (5)
Spannung	EKG-Signal	EKG-Signal (24)
Leistungsdichte	Hörschwelle	Hörschwelle (22)
Druck	Blutdruck	—
Impedanz	Hautimpedanz (Hautwiderstand)	Hautimpedanz (21)

Pr.Buch, Biostatistik, Tabelle 3

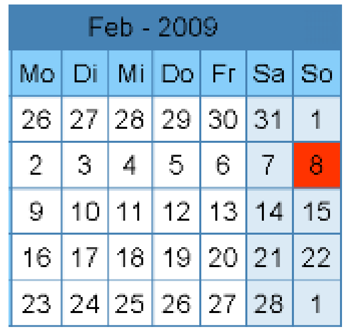
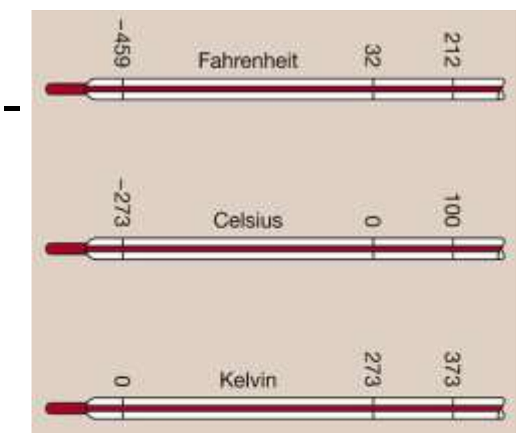


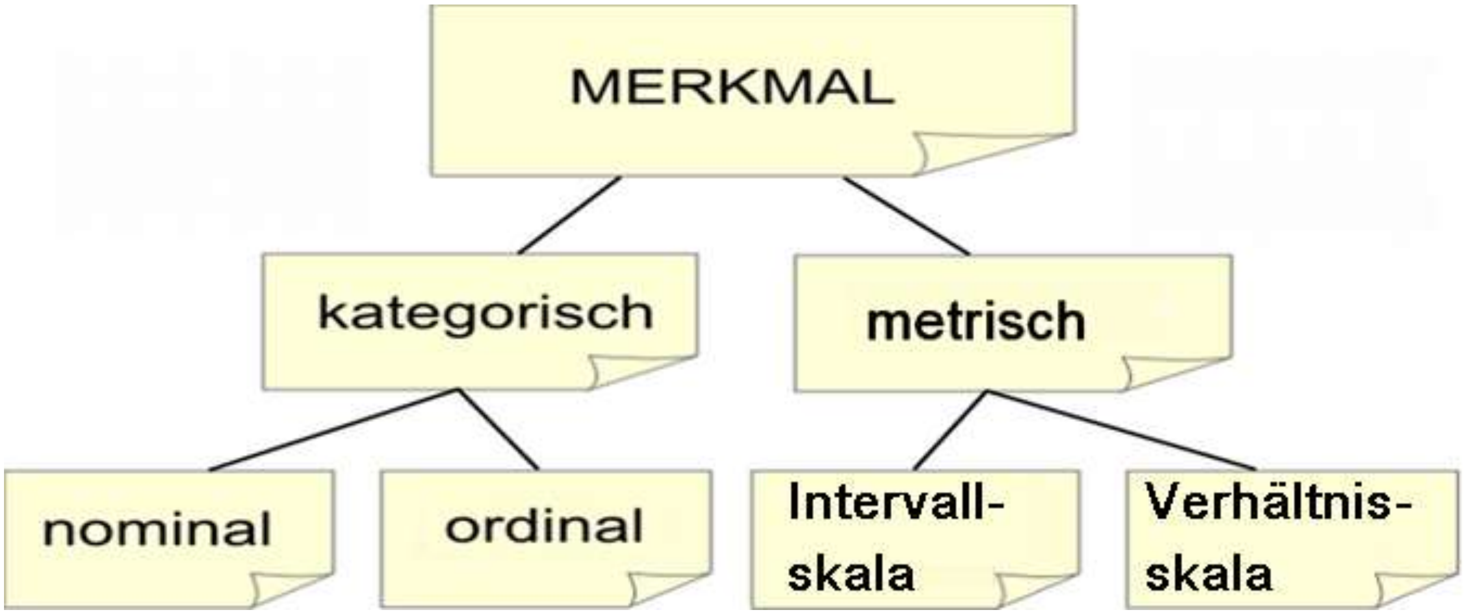
Labormessergebnisse

Name	Einheit	04.11.2004	05.10.2004	04.08.2004	05.07.2004	Min	Max
%Hypo	%		0.5		0.5	0.0	5.0
B. BURGDORFERI-AK (EIA) IGM		positiv	positiv	positiv	positiv		
B. BURGDORFERI-AK IGG (EIA)		negativ	negativ	negativ	negativ	5	10
Ery.-Vert.-Breite	%		11.6		11.6	11.5	14.5
Erythrozyten	Mill/ul	4,12	3,95	4		4	6
Haematokrit	V %		36.2	36	36.2	37.0	52.0
Haemoglobin	g/dl		12.3		12.3	12.0	16.0
Leukozyten	/ul		7		6.5	4.0	10.0
MCH	pg		32.1		32.1	27.0	34.0
MCHC	g/dl		34.0		34.0	31.0	37.0
MCV	ucm		94.4		94.4	80.0	99.0
P 18 (p18-Protein)		negativ	negativ	negativ	negativ		

- MERKMAL
  - kategorisch
    - nominal
    - ordinal
  - metrisch
    - diskret
    - kontinu-ierlich

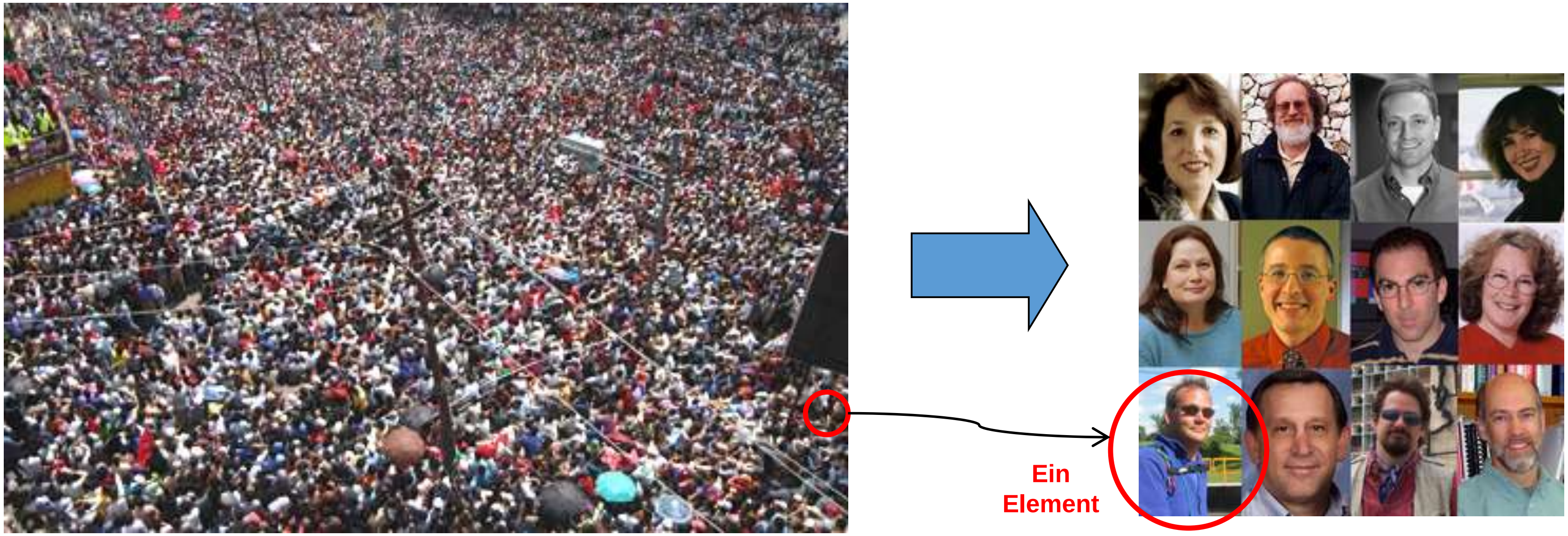
Skalentypen der metrischen Merkmale

	diskret	**kontinuierlich**
Intervall-skala definierte Differenz, „kein” 0 Punkt	Tage in einem Kalender 	Tempe-ratur in °C 
Verhältnis-skala definiertes Verhältnis, 0 Punkt	Anzahl der Zähne 	Tempe-ratur in K 



nominal	ordinal	Intervallskala	Verhältnisskala	
$=, \neq$	$=, \neq$	$=, \neq$	$=, \neq$	**Auseinander-halten**
	$<, >$	$<, >$	$<, >$	**Anordnung**
		$+, -$	$+, -$	**Differenz**
			$*, /$	**Verhältnis** 14

„Enwicklungsstand”



Stichprobe:

Grundgesamtheit (Population):

Gesamtheit der Individuen (Elemente), deren Eigenschaften bei der Studie untersucht werden sollen.

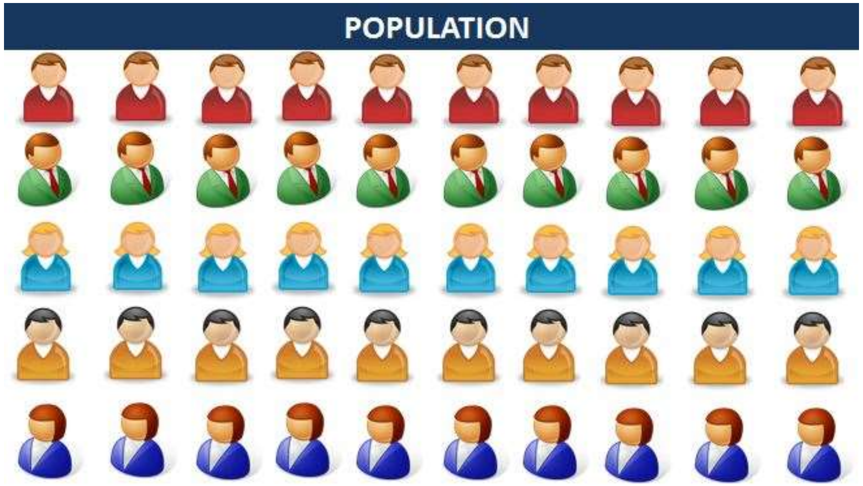
$N = \infty$ oder ungeheuer groβ

Der für die Studie ausgewählte Teil der Population.

$n << N$

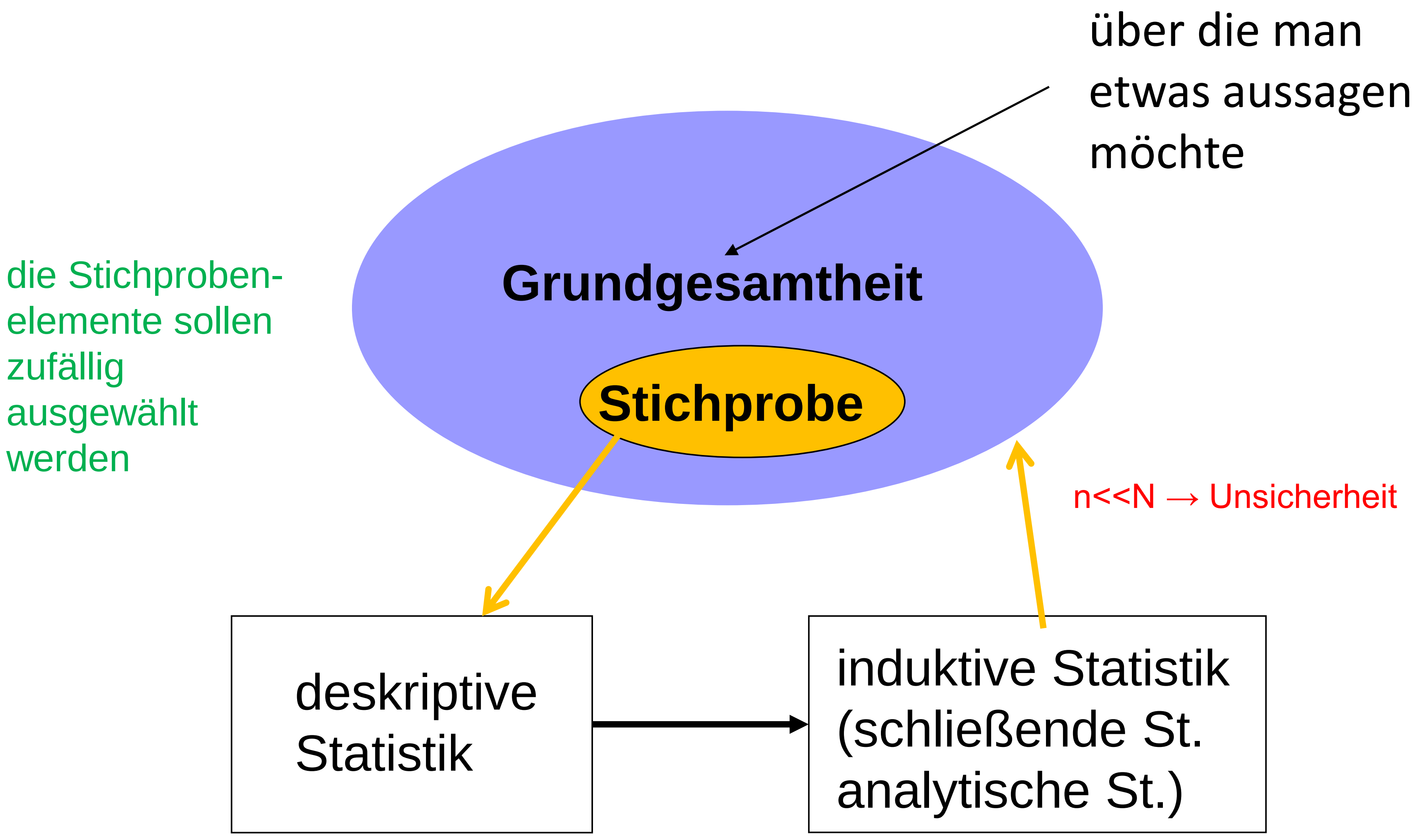
Umfang d. Stichprobe = Anzahl d. Daten

Wir brauchen eine **repräsentative Stichprobe**



Gut, repräsentativ

Nicht gut!
ganz andere Zusammensetzung als die der Population


über die man
etwas aussagen
möchte
die Stichproben-
elemente sollen
zufällig
ausgewählt
werden
Grundgesamtheit
Stichprobe
n<<N → Unsicherheit
deskriptive
Statistik
induktive Statistik
(schließende St.
analytische St.)

Frage: Wie hoch ist die normale Pulsfrequenz?

Merkmal: Pulsfrequenz (1/Min), metrisch mit Verhältnisskala

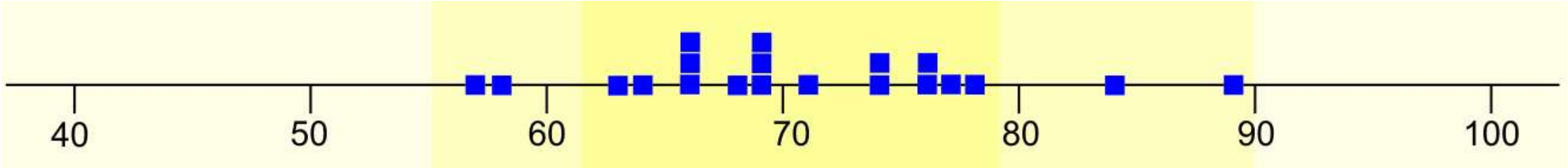
Stichprobe

66	56	89	63	66	69	71	68	58	69
78	66	64	84	74	76	69	77	74	76

Was kann man damit anfangen? (wären z.B. 700 Daten....)

Die Werte sollen **geordnet** und **verdichtet** werden.

Stellen wir die Daten entlang einer Zahlengeraden dar!



keine Daten wenige Daten viele Daten wenige Daten keine Daten

Pr.Buch Abb. 4

benutzen wir Klassen!

Unterteilen wir die Zahlengerade in gleich breite Klassen (Intervalle) und zählen wir ab, wie viele Daten sich in den so erhaltenen **Klassen** befinden!

Die Klassengrenzen sind nach Belieben festlegbar.

KLASSENGRENZEN	HÄUFIGKEIT
$55 \leq x_i < 60$	2
$60 \leq x_i < 65$	2
$65 \leq x_i < 70$	7
$70 \leq x_i < 75$	3
$75 \leq x_i < 80$	4
$80 \leq x_i < 85$	1
$85 \leq x_i < 90$	1
insgesamt:	$n = 20$

Excel:

=frequency(...)

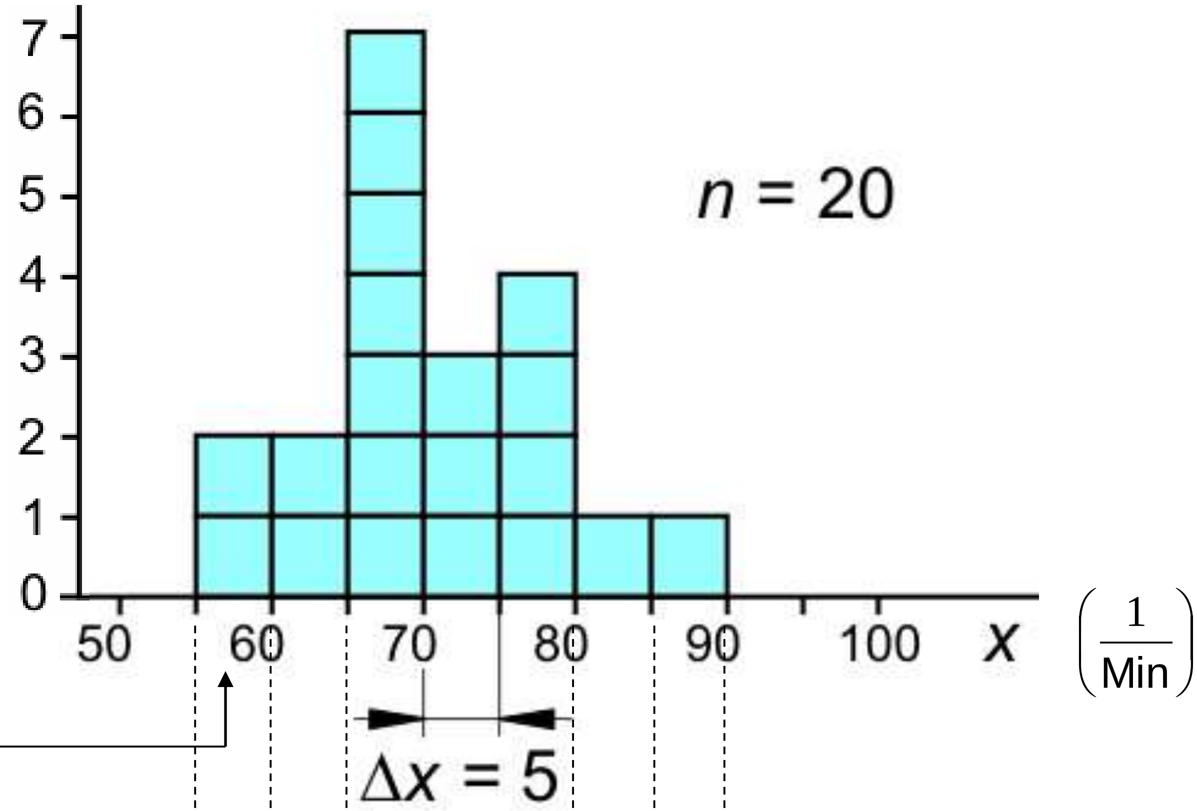
=Häufigkeit(...)

Hier z.B. Die Klassenbreite ist 5, Grenzen sind zu Zehner angepasst.

Pr.Buch Tabelle 5

Häufigkeitsdichte

$$\frac{\Delta n}{\Delta x}$$



Einheit: $\left(\frac{Stück}{5\frac{1}{\text{Min}}}\right) = \left(\frac{\text{St.} * \text{Min}}{5}\right)$

n.B. „Stück" als Einheit lässt man oft weg.

Die Fläche unter der Treppenfunktion zwischen 55 und 60:

$$5\frac{1}{\text{Min}} \cdot 2\frac{\text{Min}}{5} = 2$$

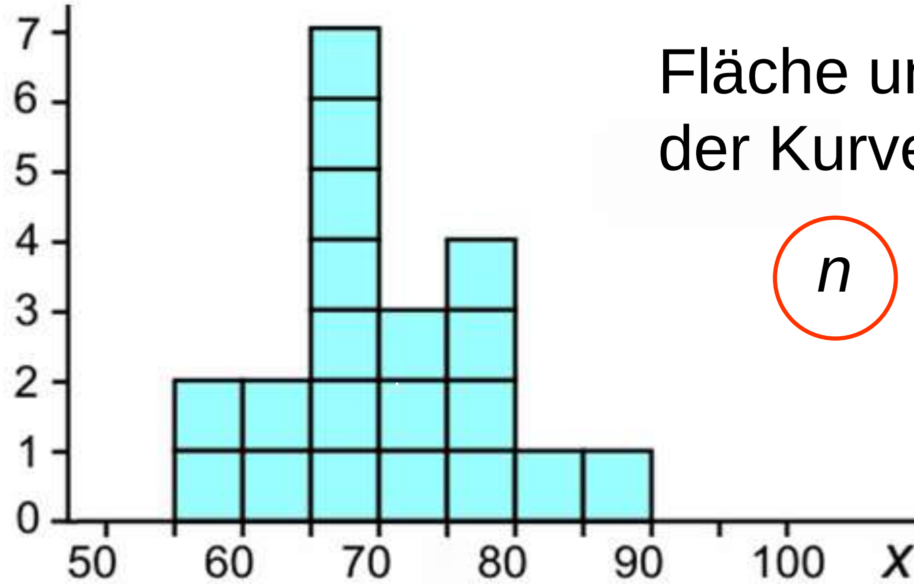
KLASSENGRENZEN	HÄUFIGKEIT
$55 \le x_i < 60$	2
$60 \le x_i < 65$	2
$65 \le x_i < 70$	7
$70 \le x_i < 75$	3
$75 \le x_i < 80$	4
$80 \le x_i < 85$	1
$85 \le x_i < 90$	1
insgesamt:	$n = 20$

Die Gesamtfläche unter der Treppenfunktion: 20 = *n*,

Anzahl der Messdaten in der Stichprobe

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

absolute



$\left(\frac{1}{\text{Min}}\right)$

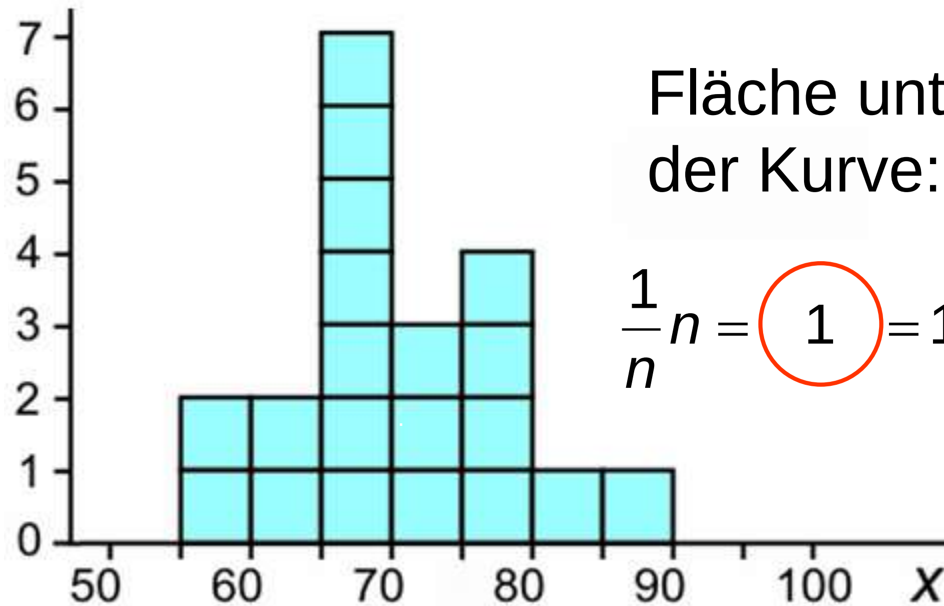
Fläche unter der Kurve:

n

Häufigkeitsdichte-verteilung

$\frac{1}{n}\frac{\Delta n}{\Delta x} \left(\frac{1}{20}\frac{\text{Min}}{5}\right)$

relative

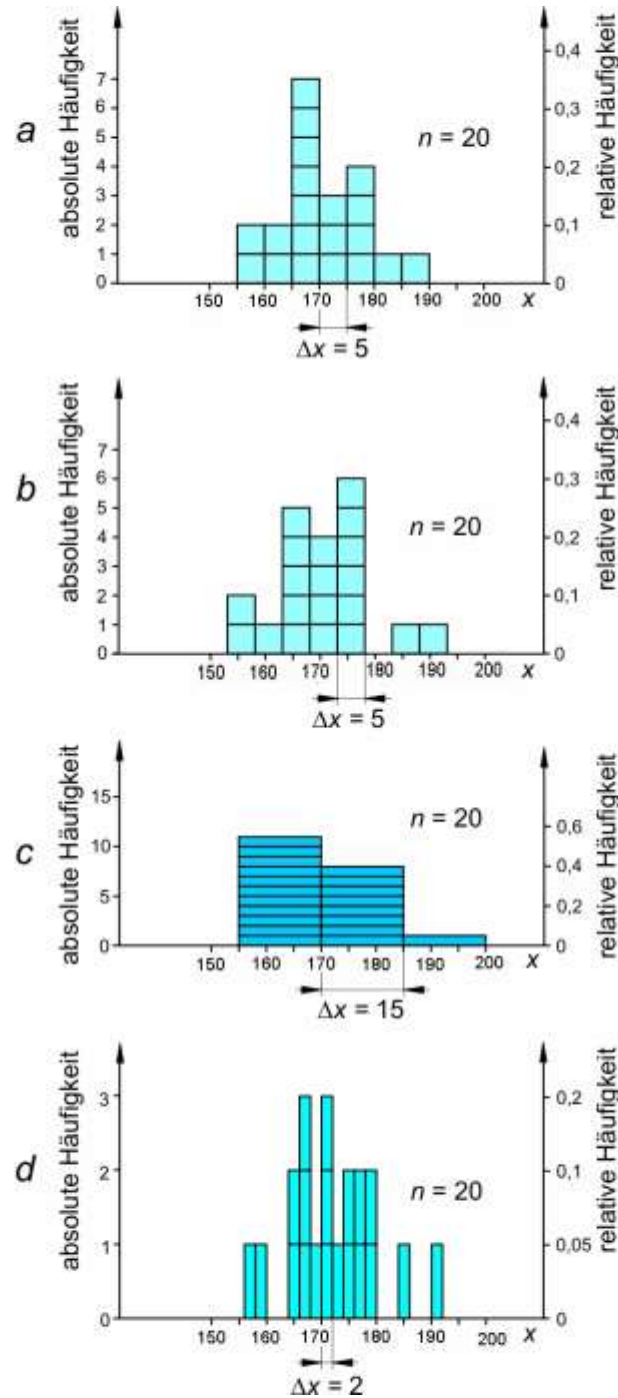


$\left(\frac{1}{\text{Min}}\right)$

Fläche unter der Kurve:

$\frac{1}{n} n = 1 = 100\,\%$

Die Klassenbreite kann das Aussehen des Histogramms wesentlich beeinflussen, wenn die Datenmenge nicht groß genug ist.
In diesem Fall gibt es auch eine relativ hohe Instabilität des Histogramms



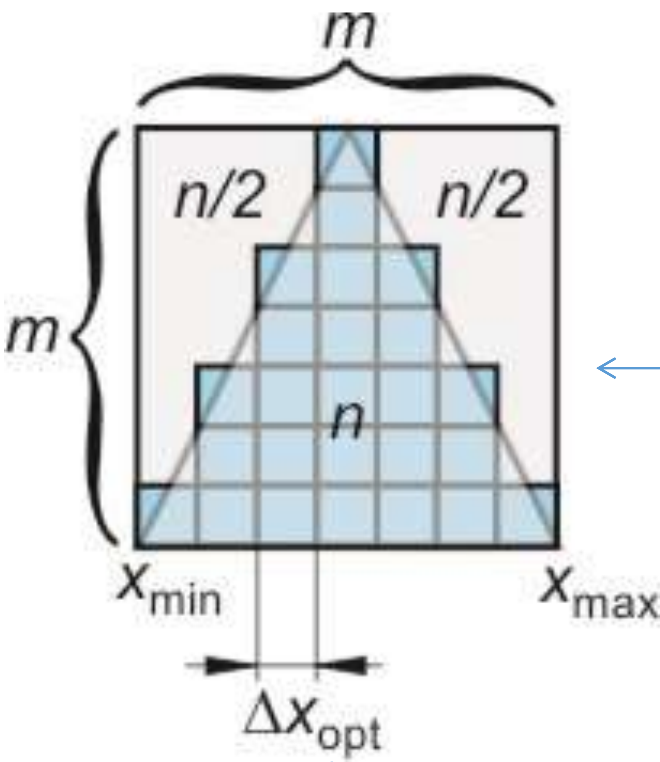
Selbe Grundgesamtheit, 2 Stchproben

Zu große Klassenbreite

Zu kleine Klassenbreite

Pr.Buch Abb. 5

Bestimmung der optimalen Klasseneinteilung



Weil oft die Daten um einem zentralen Wert gestreut sind, hat das Histogramm ein „Gipfel".

optimale Klassenanzahl *m Stück*:

$$m^2 = 2n$$

$$m = \sqrt{2n} \qquad m = \sqrt{40} = 6.3$$

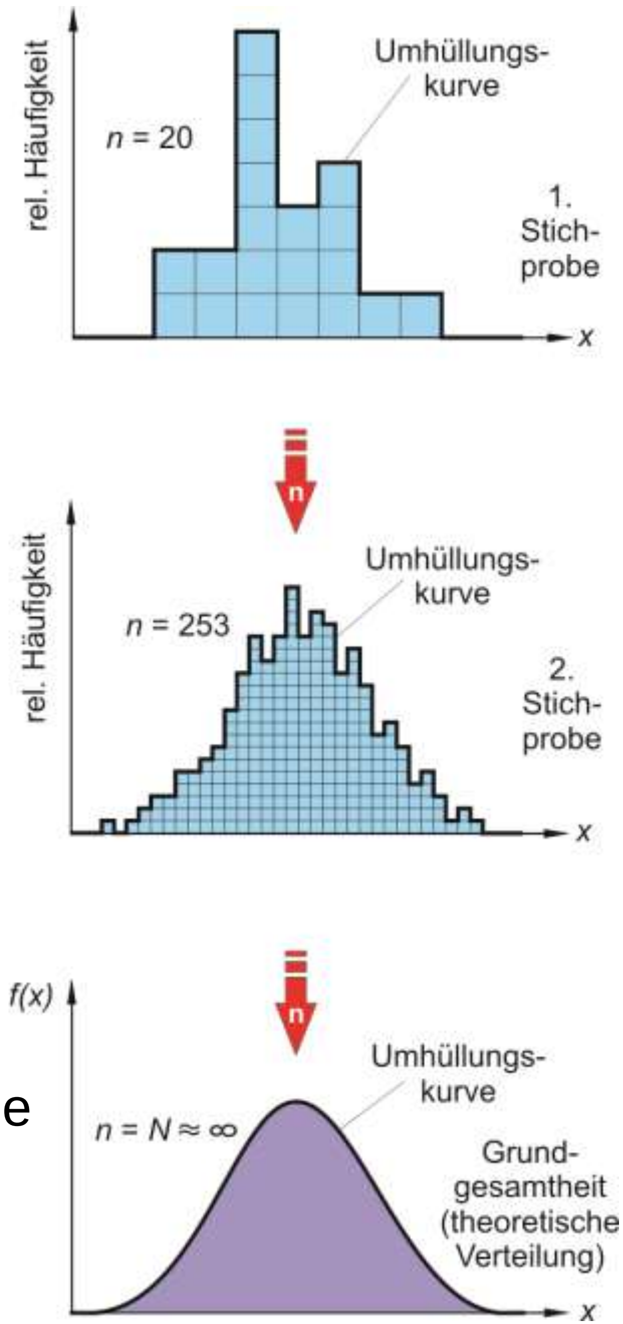
optimale Klassenbreite Δx:

$$\Delta x = \frac{x_{max} - x_{min}}{m} \qquad \Delta x = \frac{89 - 56}{6.3} = 5.2$$

empirische Funktion

empirische Funktion

theoretische Funktion



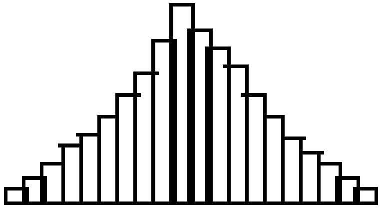
Pr.Buch Abb. 6

n vergrößert sich,
die Klassenbreite Δx kann verkleinert werden

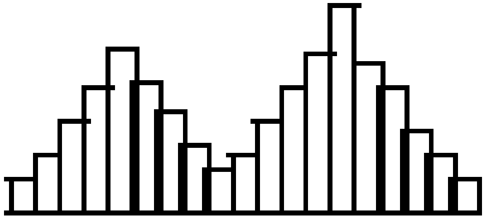
Bei großen Stichproben ergibt die empirische Verteilungsfunktion **eine sehr gute Näherung** der theoretischen Verteilungsfunktion. (Die Stichprobe ist „fast gleich” der Grundgesamtheit.)

Analyse von Häufigkeitsverteilungen

homogene symmetrische Stichprobe:

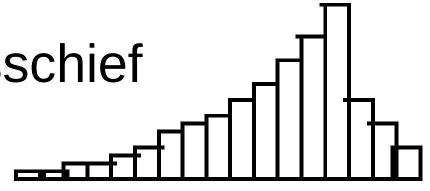


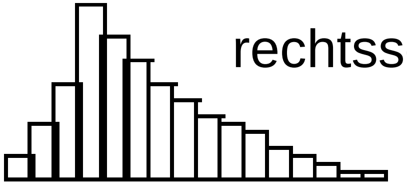
heterogene Stichprobe:



homogene nichtsymmetrische Stichproben:

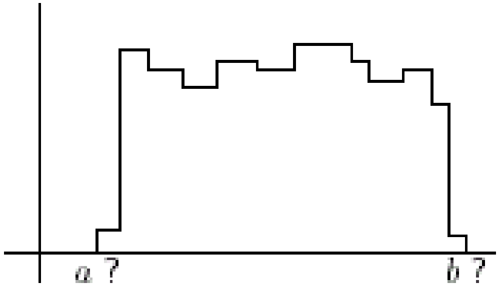
linksschief



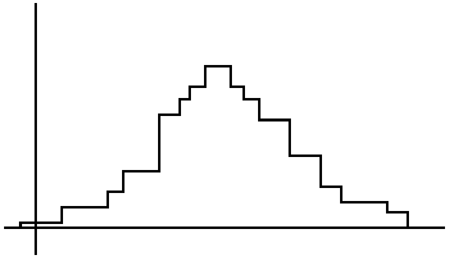


rechtsschief

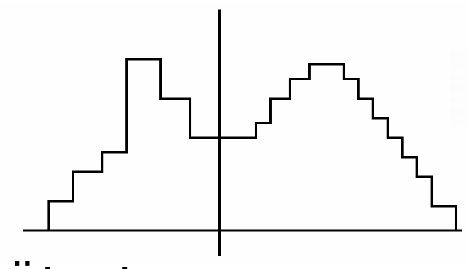
Vermutungen macht man auch:



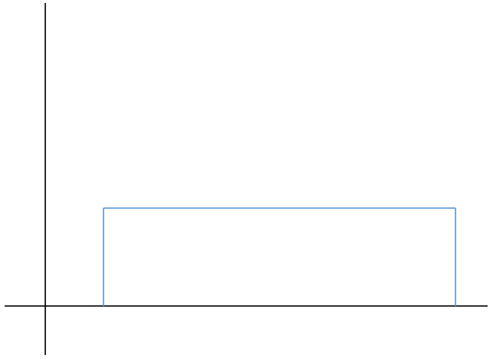
Gleichverteilung?

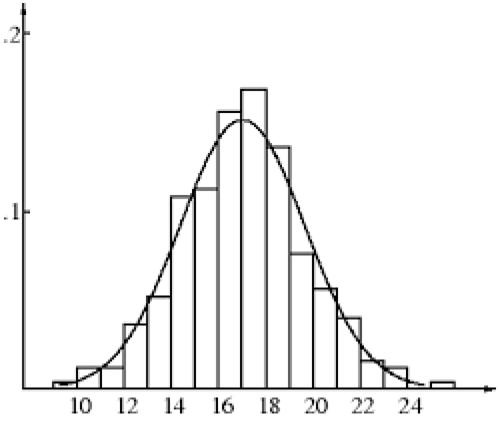


Normalverteilung?



Überlagerung von zwei Normalverteilungen?

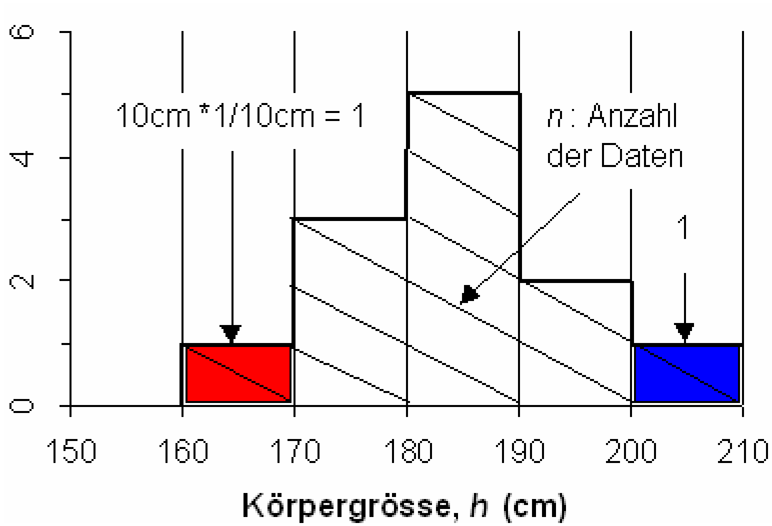




Vergleichen mit bekannten Verteilungen...

Häufigkeitsverteilung

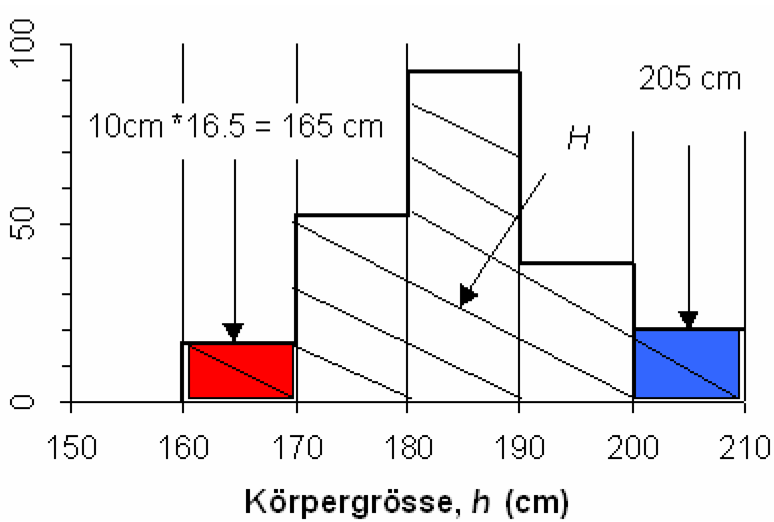
$$\frac{\Delta N}{\Delta h} \quad \left(\frac{1}{10\,\text{cm}}\right)$$

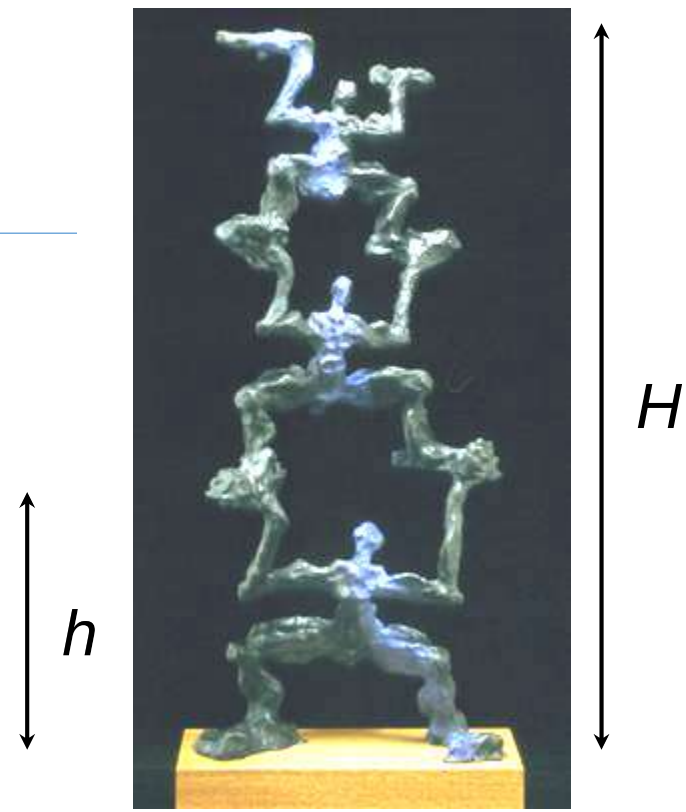


h: Körperhöhe

H: kollektive Höhe, Gesamthöhe

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

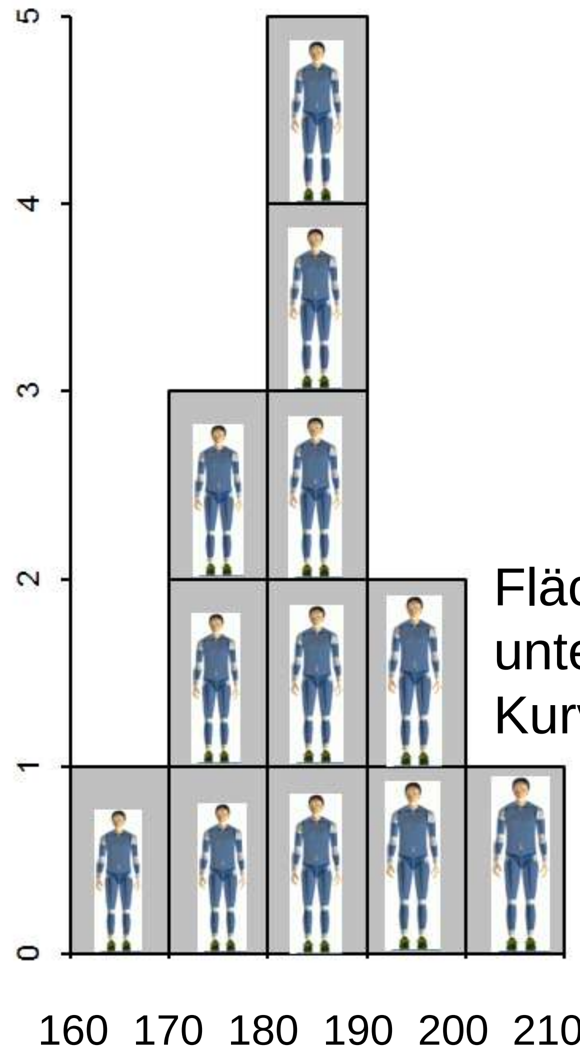




Spektrum ist eine spezielle Häufigkeitsverteilung

Häufigkeitsdichte

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



Fläche unter der Kurve: *n*

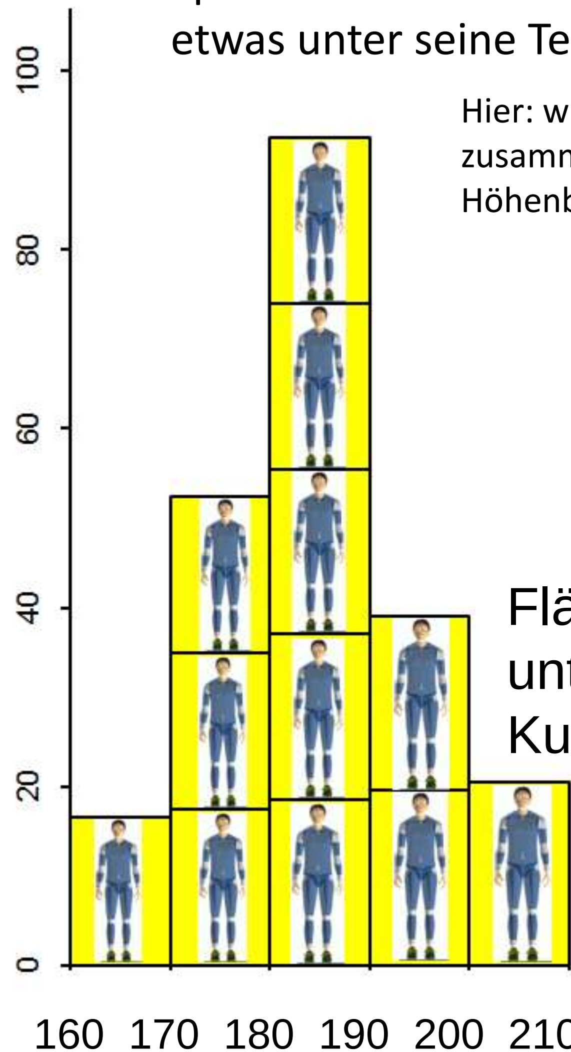
h (cm)

Spektrum

Spektrum ist die Verteilung von etwas unter seine Teile

Hier: wie viel Höhe ist zusammen in einem Höhenbereich

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



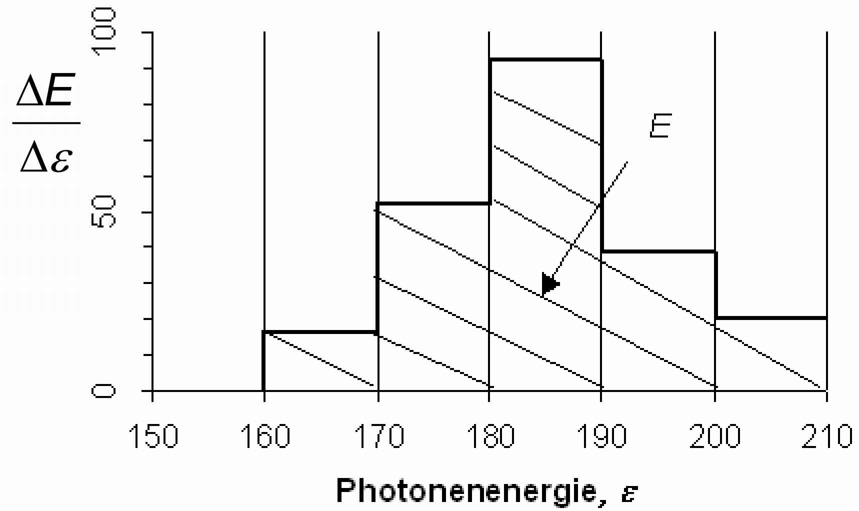
Fläche unter der Kurve: *H*

h (cm)

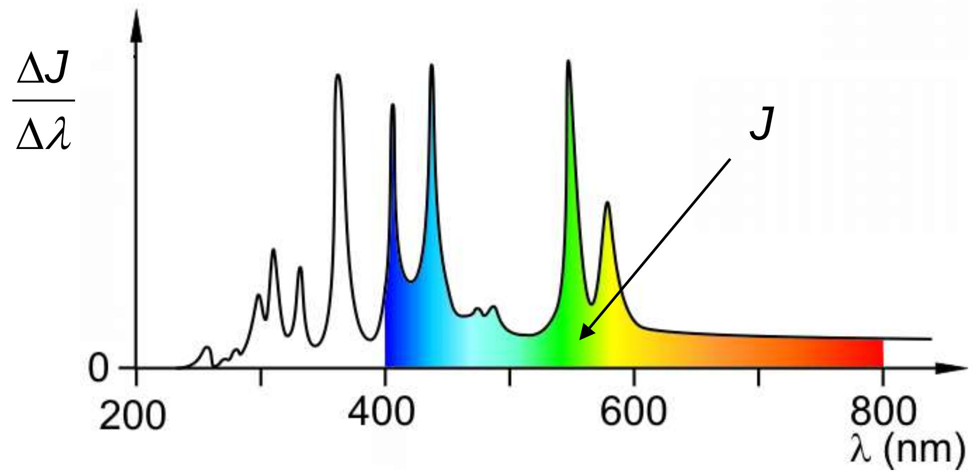
Beispiel aus der Physik

Emissionsspektrum:

wie verteilt sich die emittierte Energie über die Photonenenergien



charakteristische Größe des Energietransports: **Intensität**



Benützung der **Wellenlänge** ist bequemer als die der Photonenenergie

Lageparameter. Charakterisierung das Zentrum der Daten

Durchschnittswert (der arithmetische Mittelwert)

=average(...)

=Mittelwert(...)

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

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

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

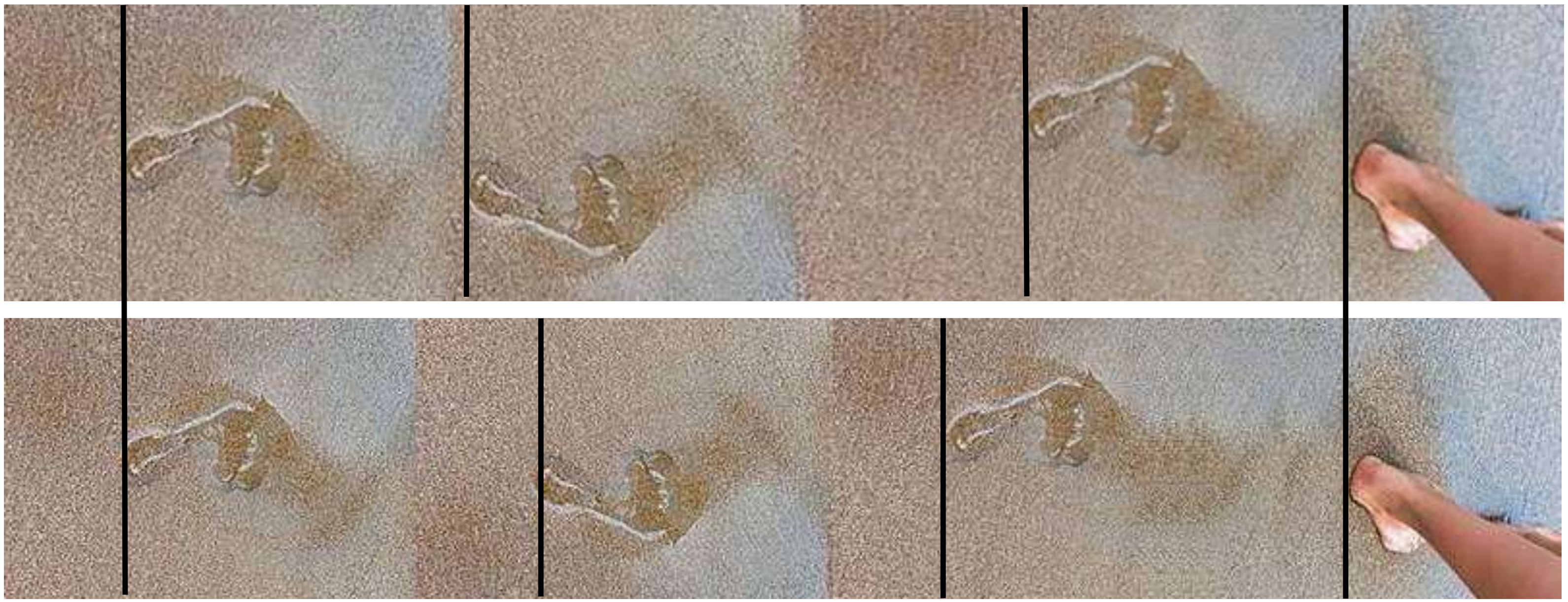
=median(...)

=Median(...)

Durchschnittswert (der arithmetische Mittelwert)

Beispiel: Schritte

$$x_1 \quad + \quad x_2 \quad + \quad x_3 \quad =$$



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

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

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

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

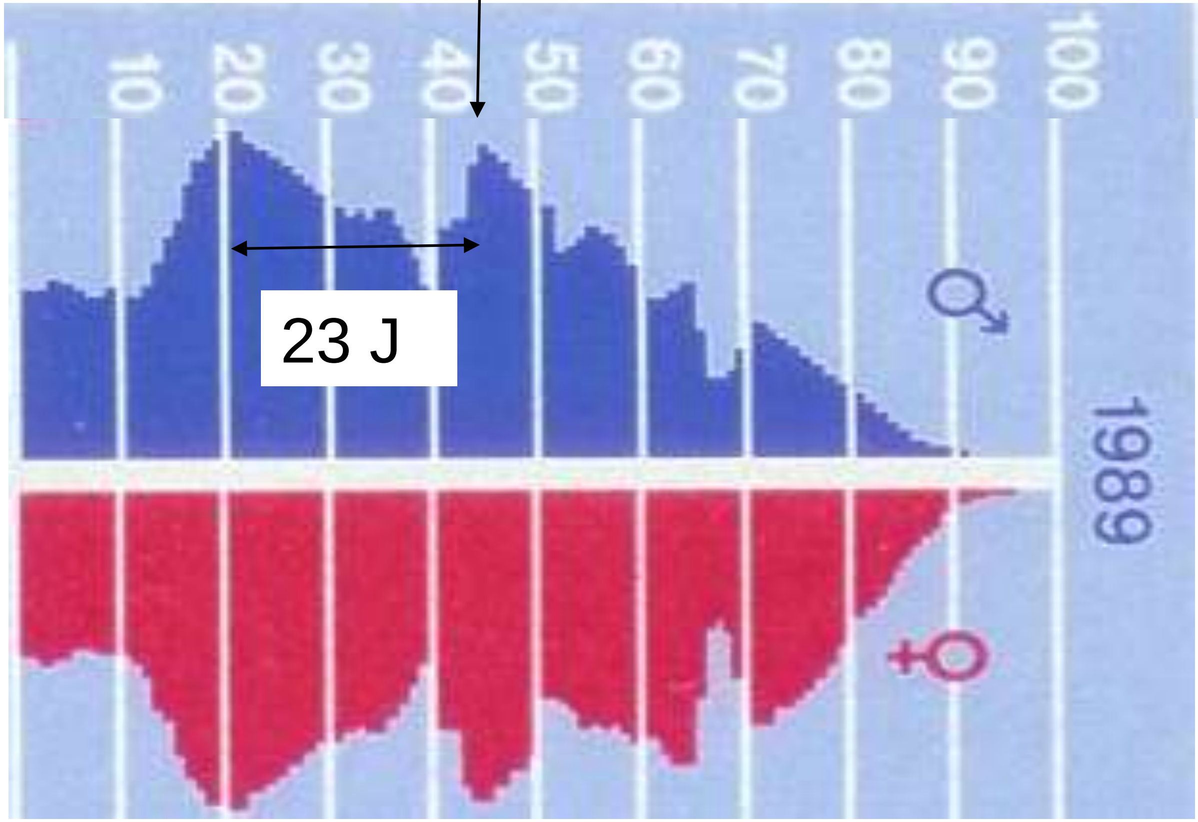
=Mittelwert(...)

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.

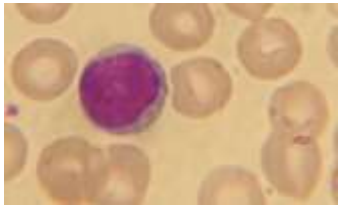
1989-43=1946

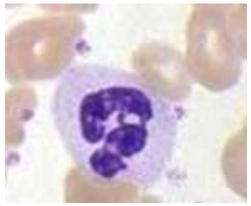


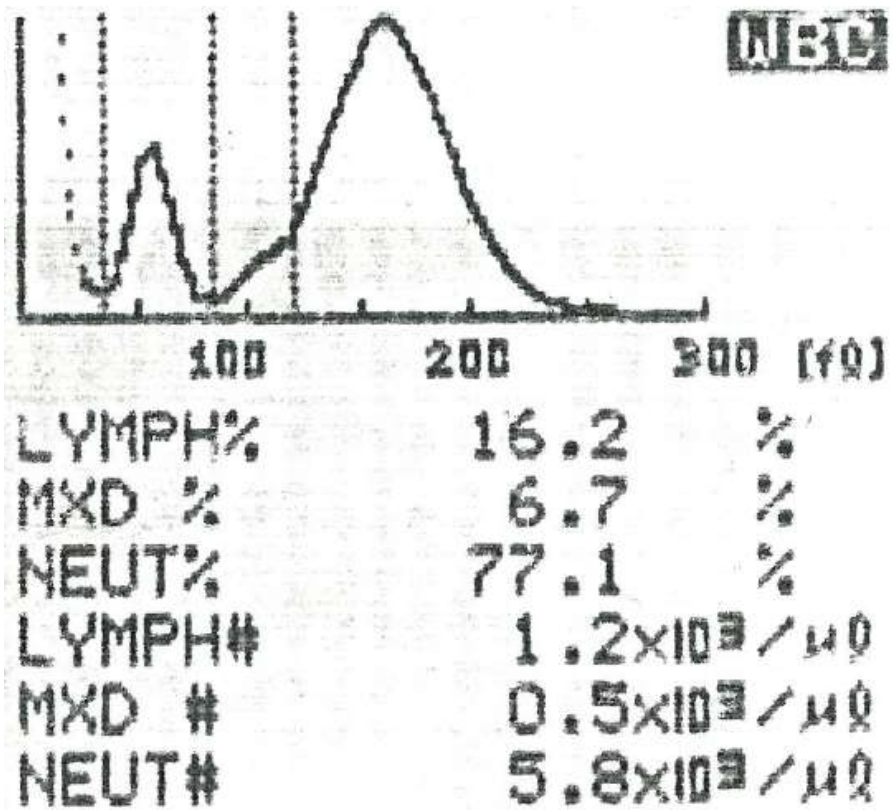
Beispiel in der Medizin

Coulter Zähler

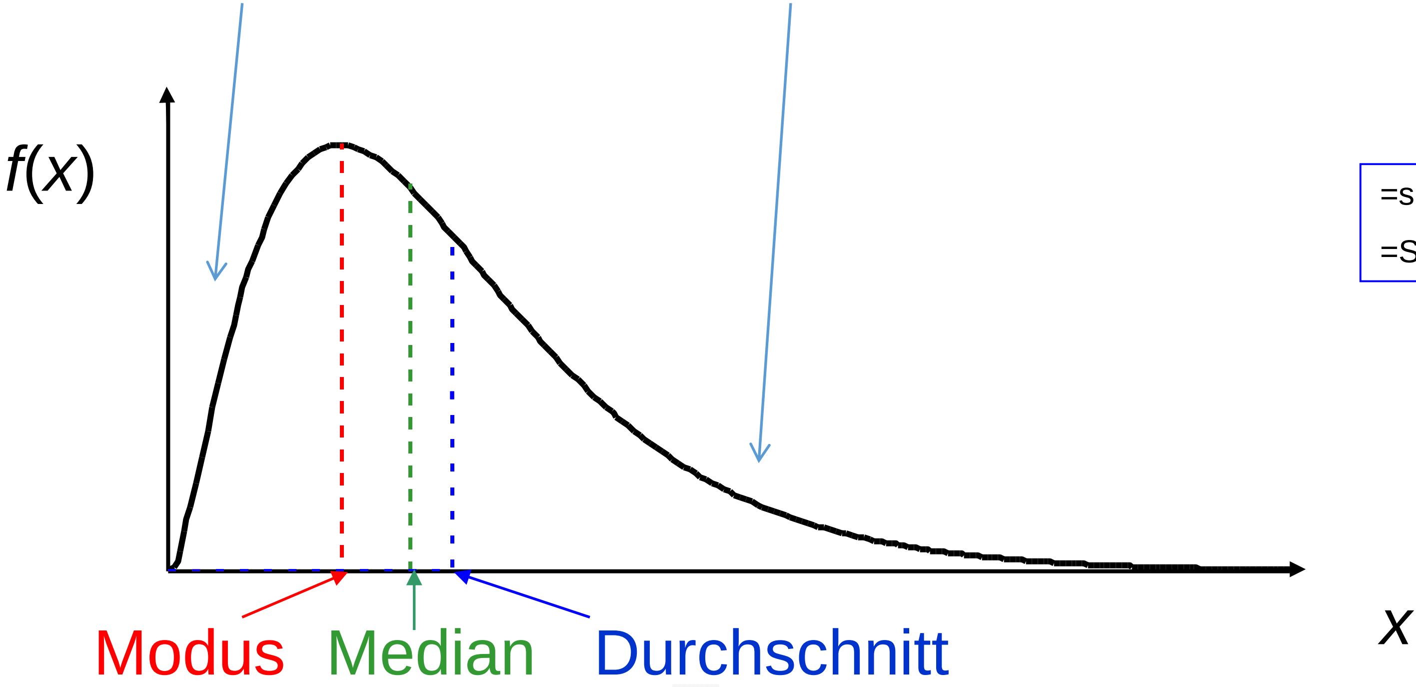
Grössenverteilung der Zellen







Linkssteile bzw. rechtsschiefe Verteilung



=skew(...) > 0

=Schiefe(...) > 0

>0 : rechts
<0 : links

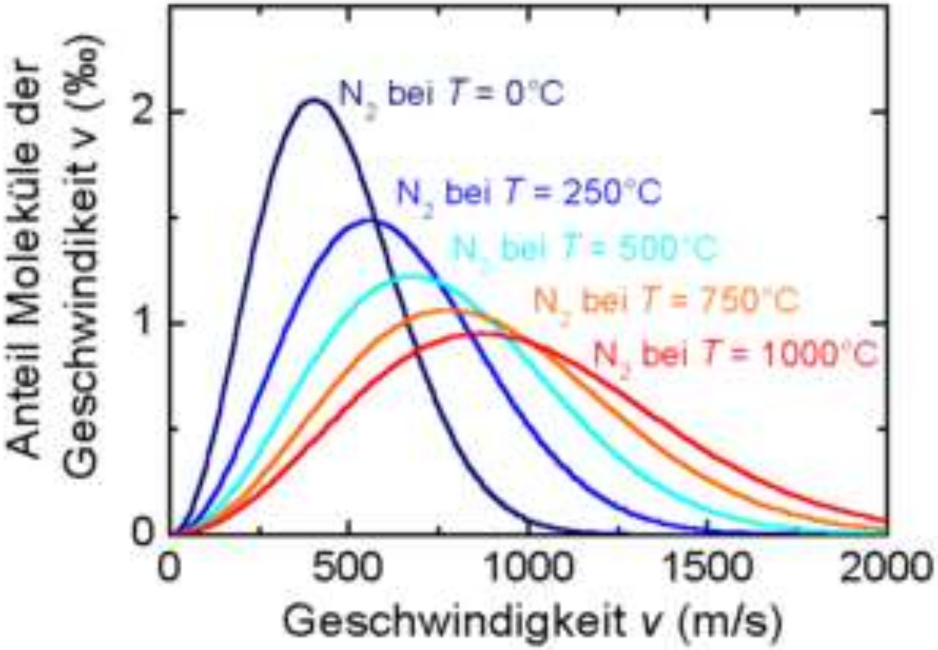
z.B. Einkommensverteilungen in einem Land:

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

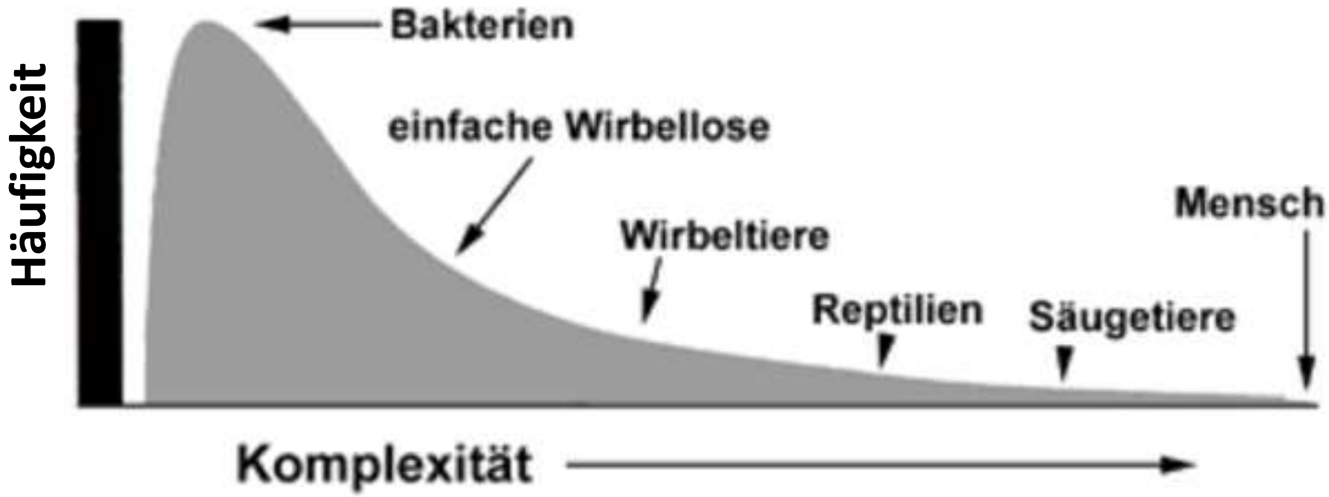
Weitere Beispiele

Maxwell-Boltzmann-Verteilung

(siehe später in der Physik)

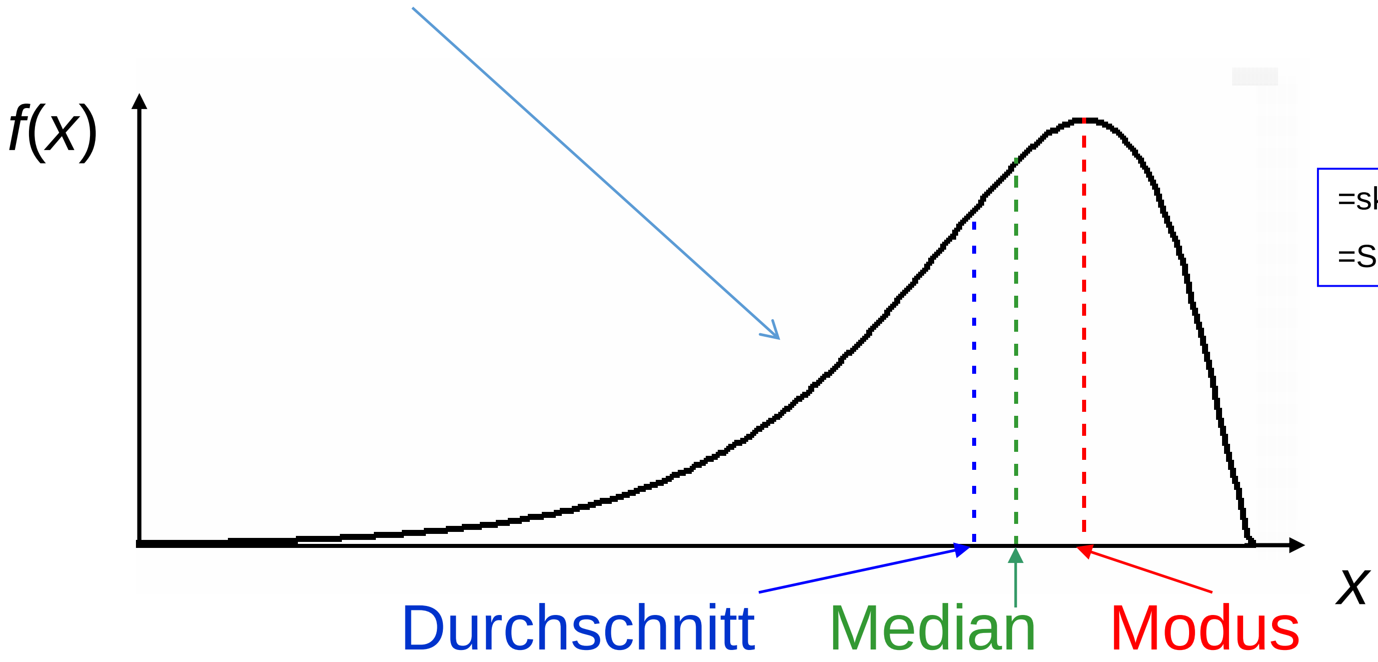


Komplexität der Tiere



www.vordenker.de/if_gould/images/verteilung.gif

Linksschiefe bzw. rechtssteile Verteilung

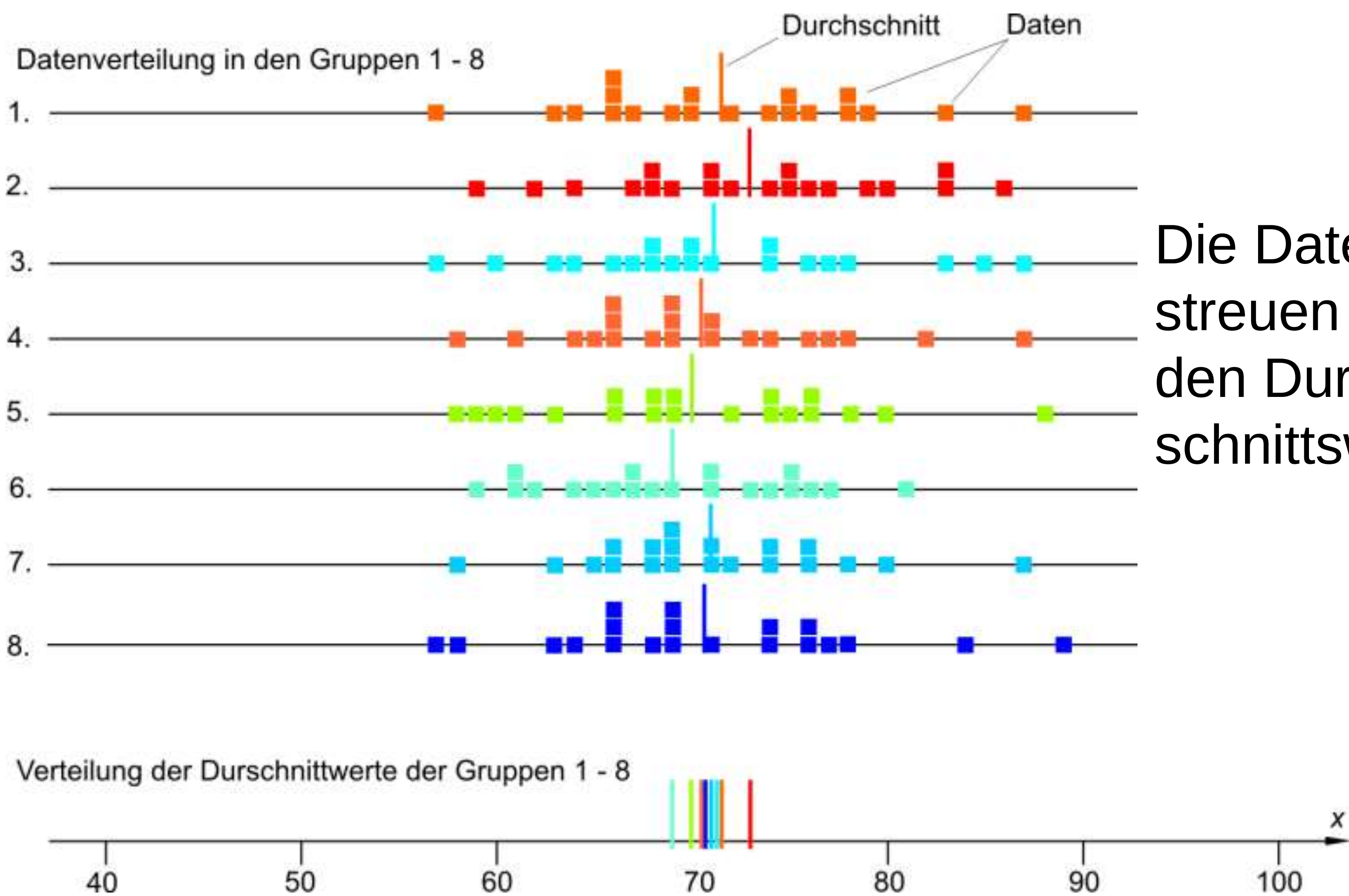


=skew(...) < 0

=Schiefe(...) < 0

z.B. Dauer einer Schwangerschaft

Daten und ihre Durchschnittswerte



Die Daten streuen um den Durchschnittswert.

Pulsfrequenzen (1/Min)

Pr.Buch Abb. 10

Streuungsparameter. Charakterisierung der Variation der Daten

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

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

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

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

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

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

Spannweite: x_{max}-x_{min}

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

α-Quantil

$0 < \alpha < 1$

(seien dazu die x_i aufsteigend sortiert):

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

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

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

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

$\boxed{\text{=Quartil(...)}}$

Hier kann nur α = einige quartile sein!

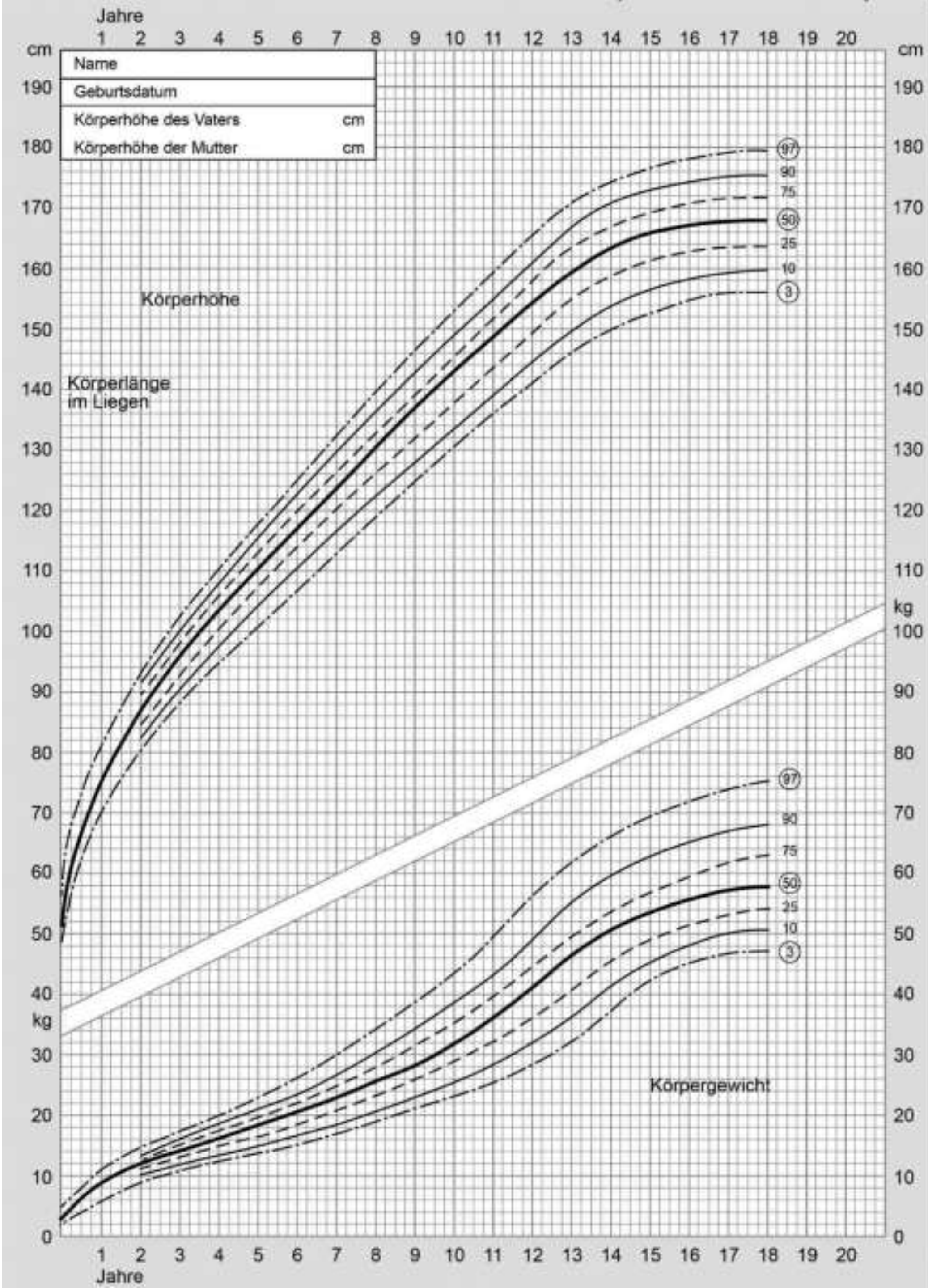
mit Wörter: z.B. **Dezile**

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

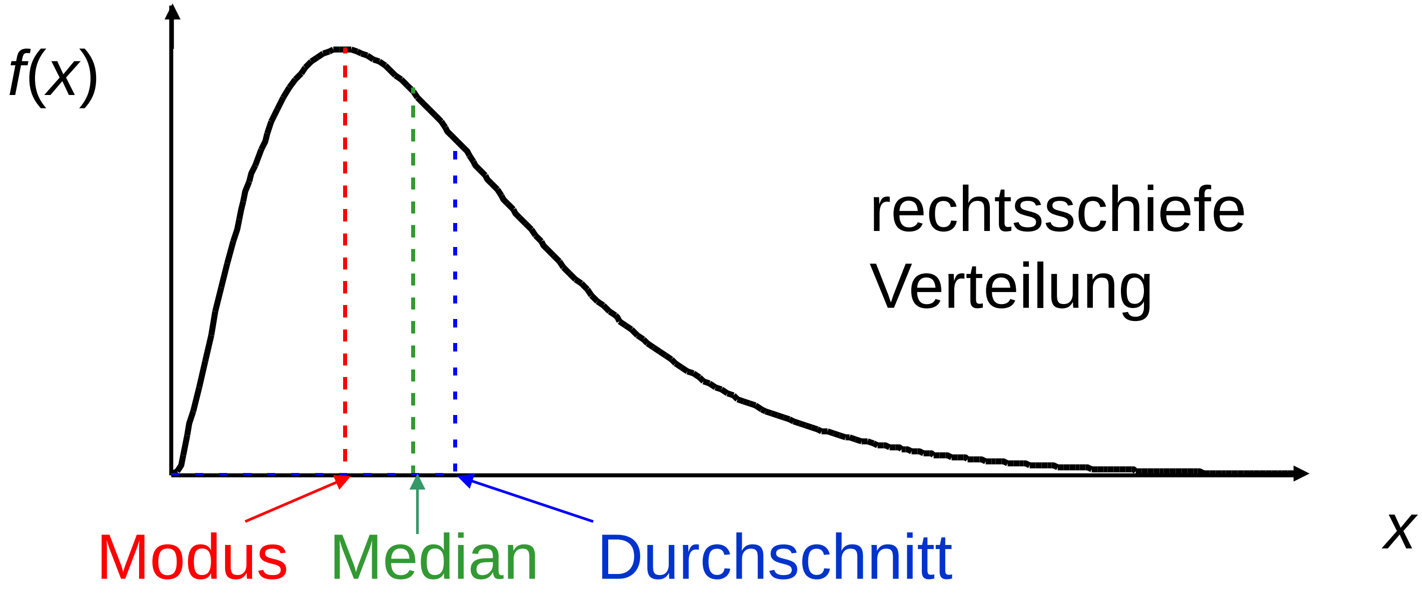
Perzentilenkurven sind ein Werkzeug für den Arzt.

Wachstums- und Gewichtskurven für Mädchen

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



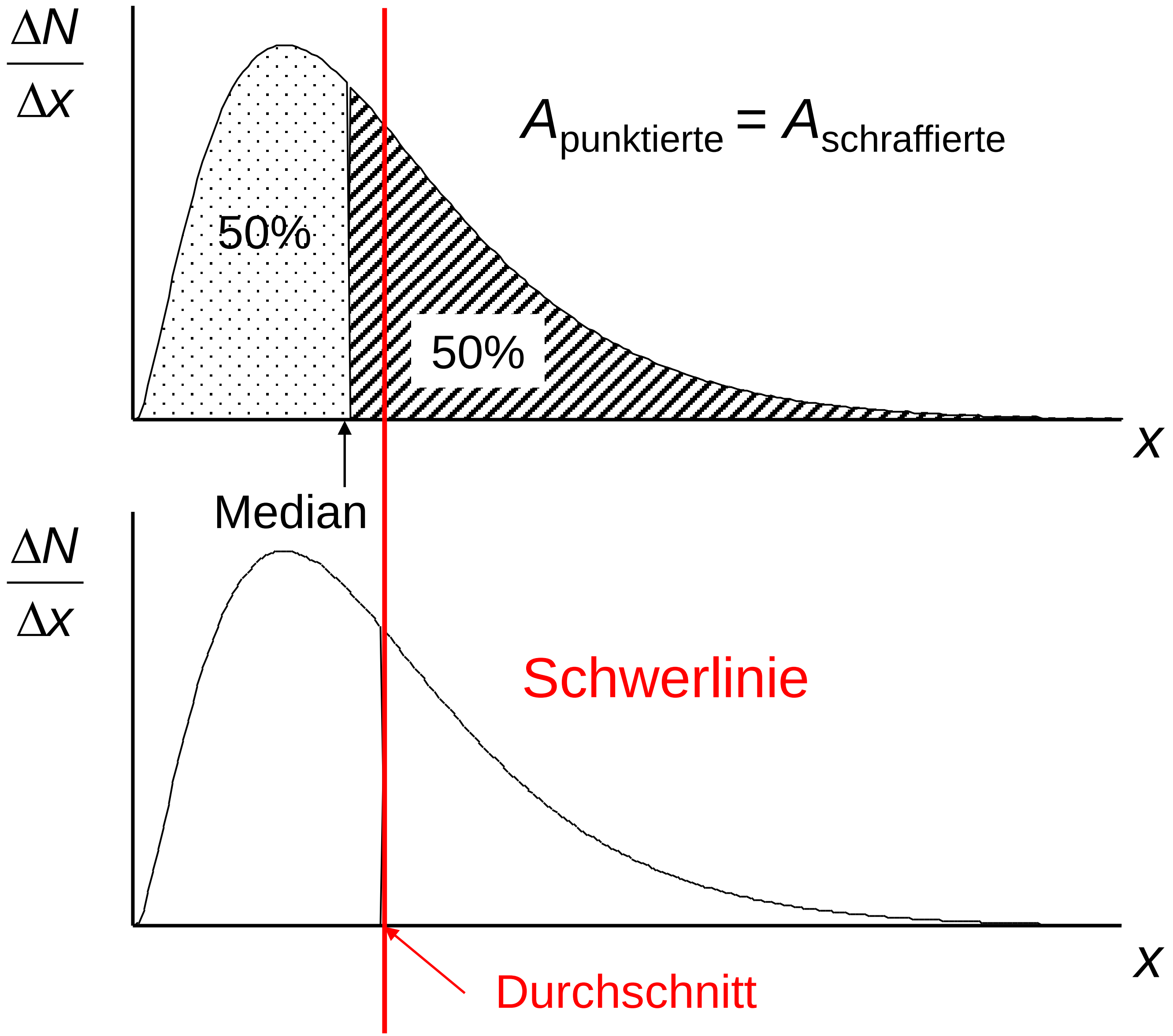
Die Lageparameter sing generell nicht identisch.



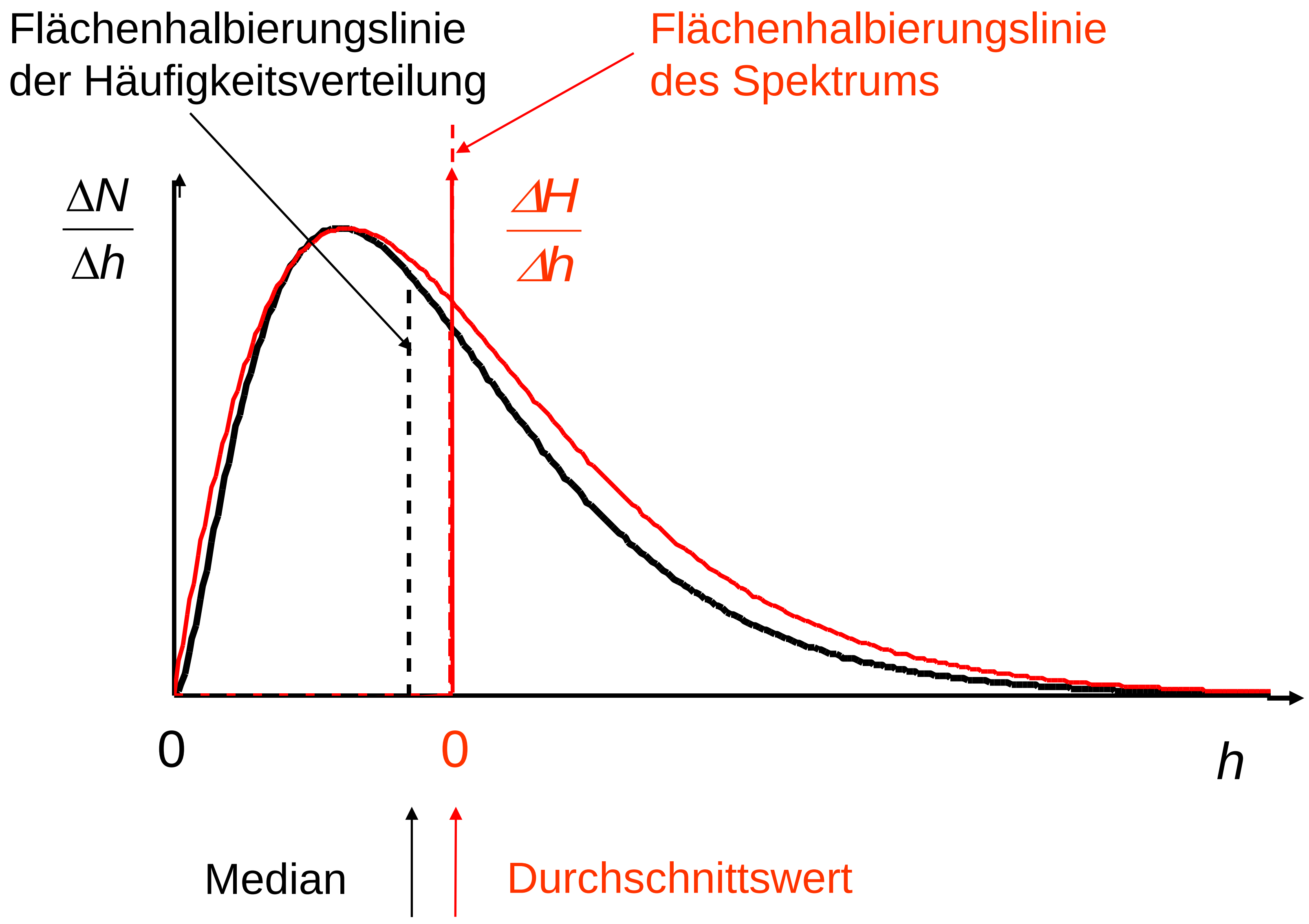
Vorsicht mit Skalentypen!
Besonders mit zahlen: die originelle Skala kann gut „nur” nominal, oder ordinal sein (z.B. Noten)

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

Position des Medians und des Durchschnitts einer Verteilung (1)

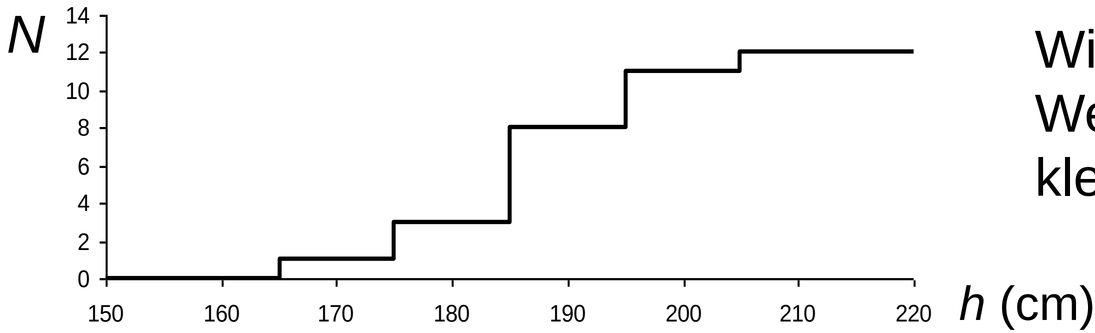


Position des Medians und des Durchschnitts einer Verteilung (2)



Summen- (kumulierte/kumulative) Häufigkeitsverteilung

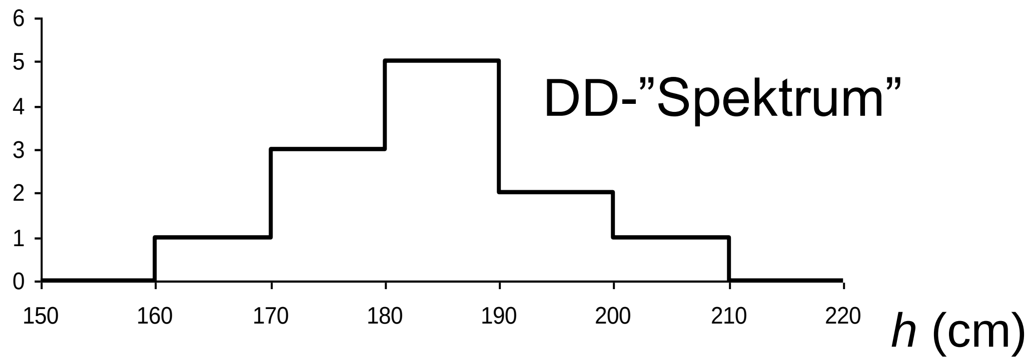
Summen-
Häufigkeits-
verteilung



Wieviele
Werte sind
kleiner als h?

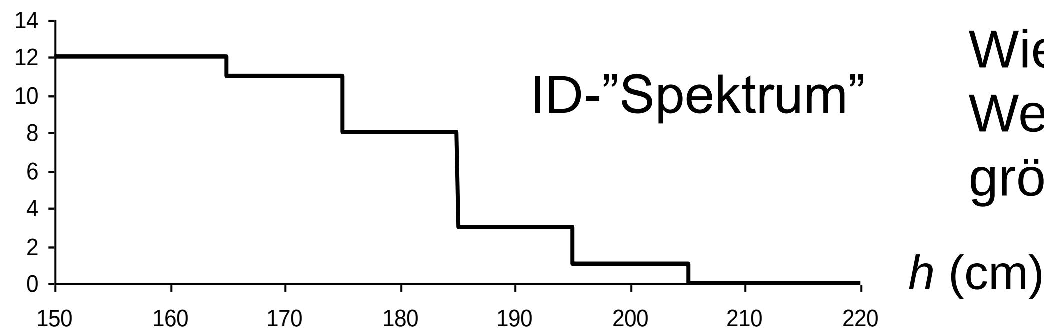
$$\frac{\Delta N}{\Delta h} \quad \left(\frac{1}{10\,\mathrm{cm}}\right)$$

Häufigkeitsdichte-
Verteilung



$M=N_0-N$

„Summen-
Häufigkeits-
verteilung"



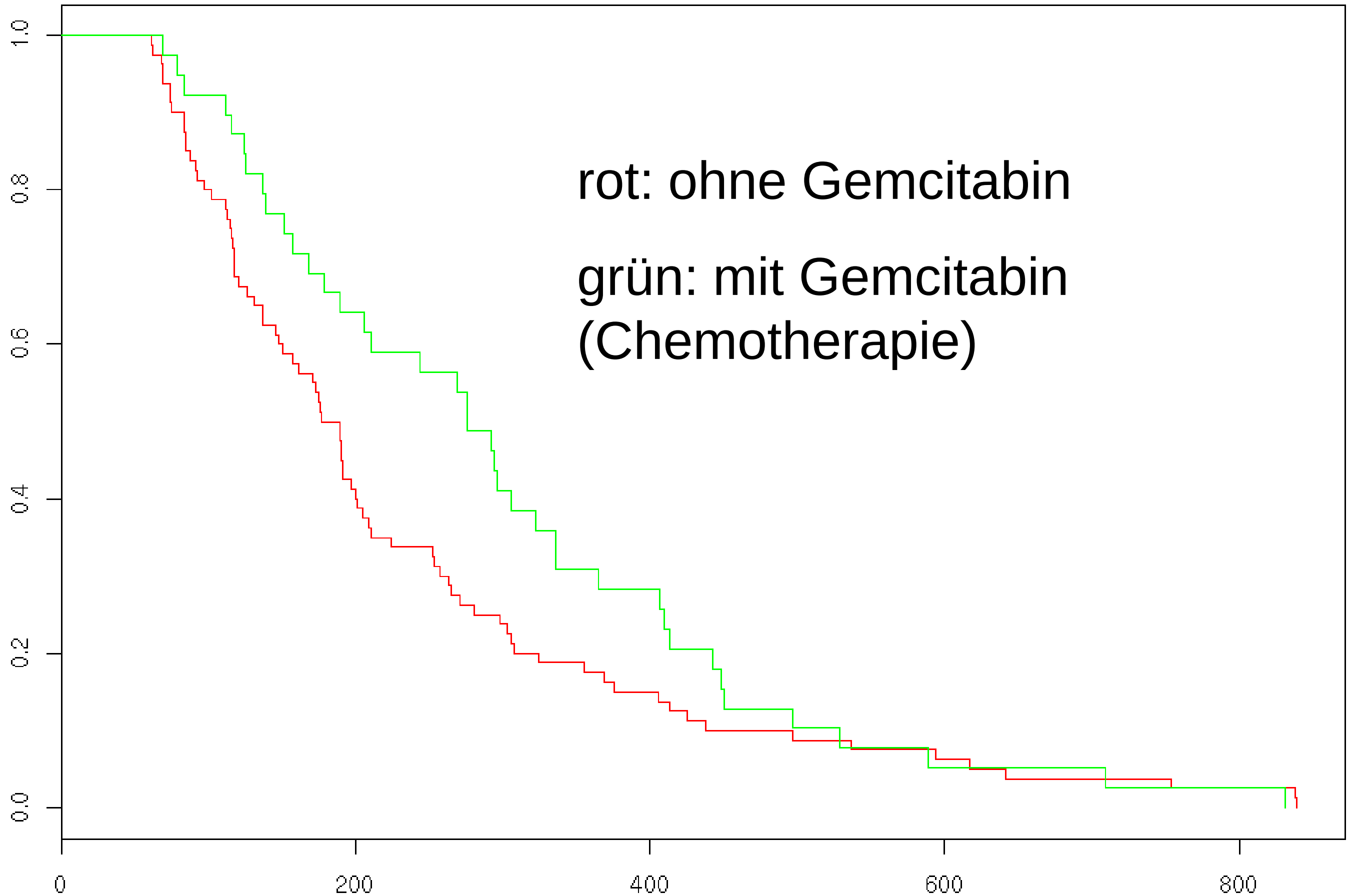
Wieviele
Werte sind
grösser als h?

Überlebenskurven

relative
„Summen-
Häufigkeits-
verteilung”

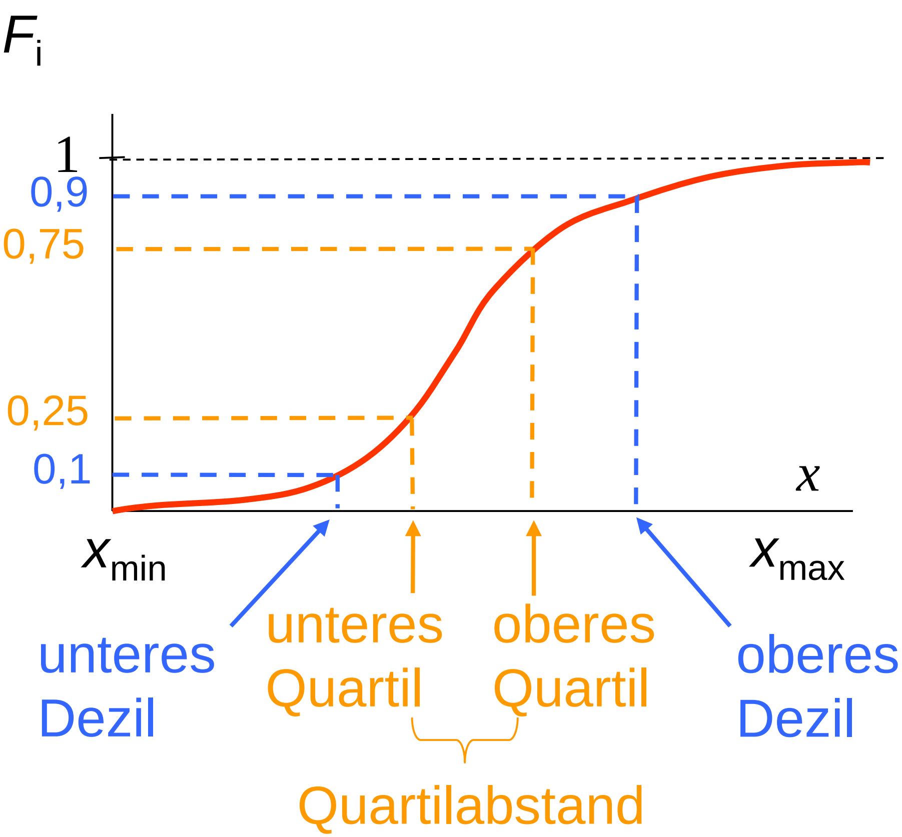
kumulatives
Überleben
nach der
Operation

Wirkung der Chemotherapie. Pankreaskarzinom



Überleben, Tage

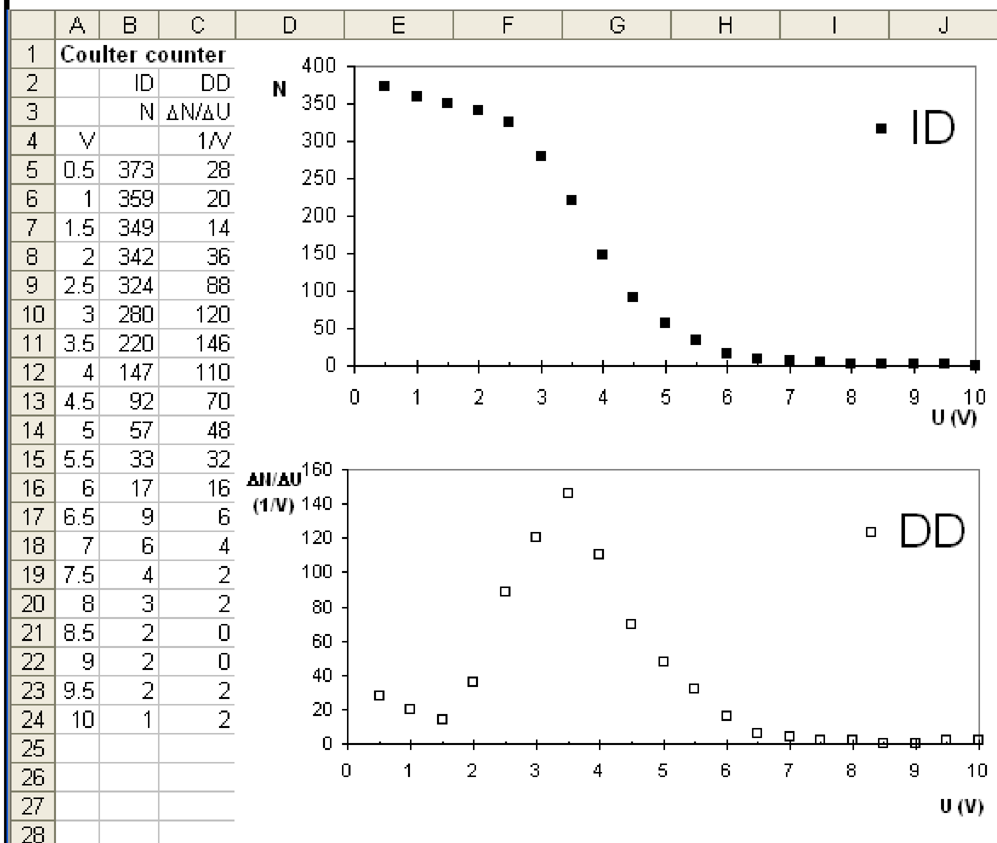
Quantile und die relative Summenhäufigkeits-verteilung



Beispiel in der Physikpraktikum:

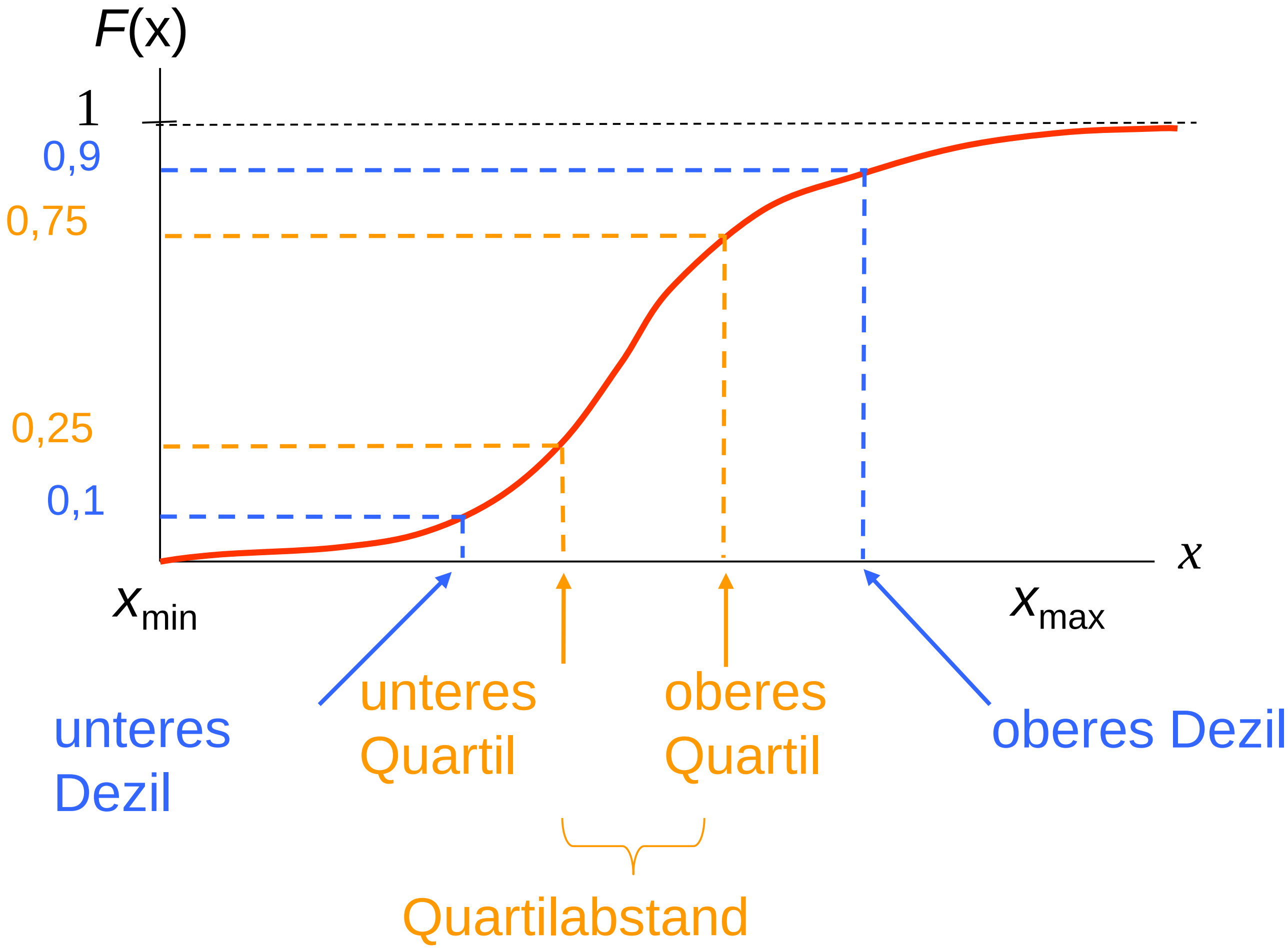
Coulter Zähler

(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
25			
26			
27			
28			

Quantile und die relative Summenhäufigkeits-verteilung

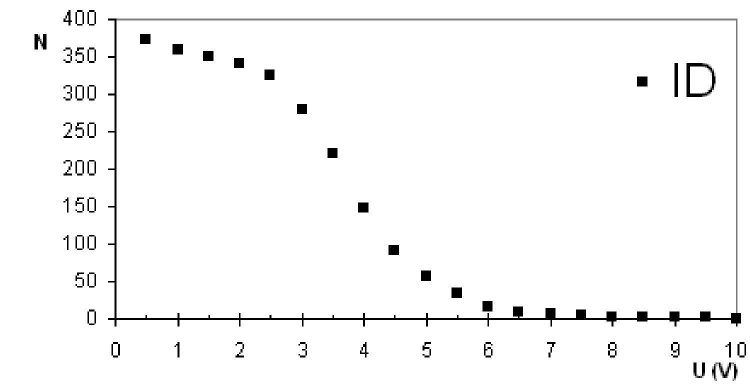


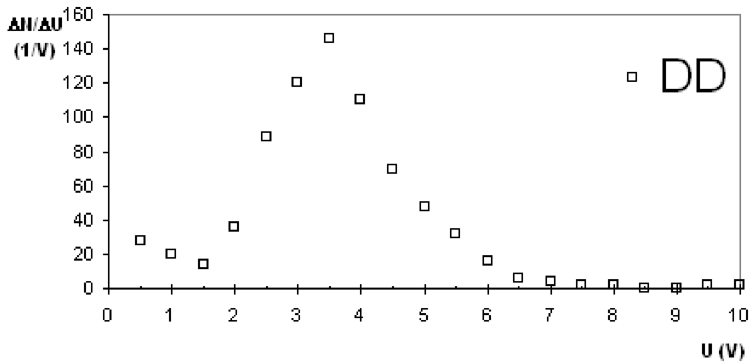
Beispiel in der Physikpraktikum:

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
25			
26			
27			
28			





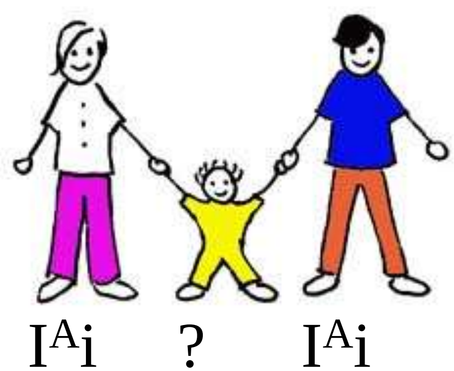
Verteilungen und Schätzungen
Grundlagen der Wahrscheinlichkeitslehre

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

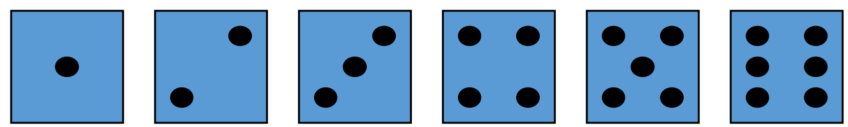
die einzelnen, nicht mehr zerlegbaren und sich gegenseitig ausschliessenden Ausgänge oder Ergebnisse eines Zufallsexperimentes

Ereignismenge, Ereignisraum (Ω)

Reihe aller möglichen Elementarereignisse. Z.B:

beim Würfelspiel:

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



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

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

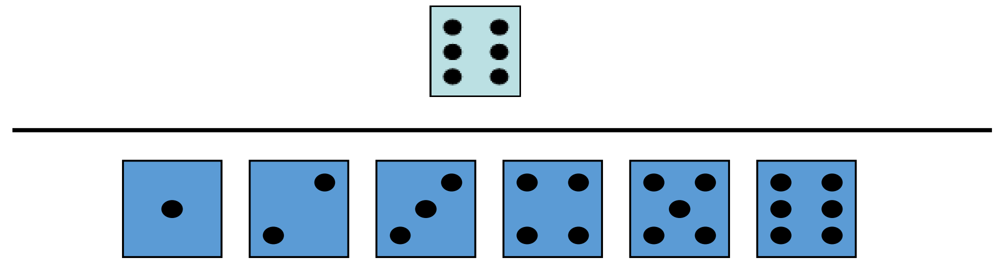
$$\boxed{p(E)=\frac{g}{m}}$$

günstig

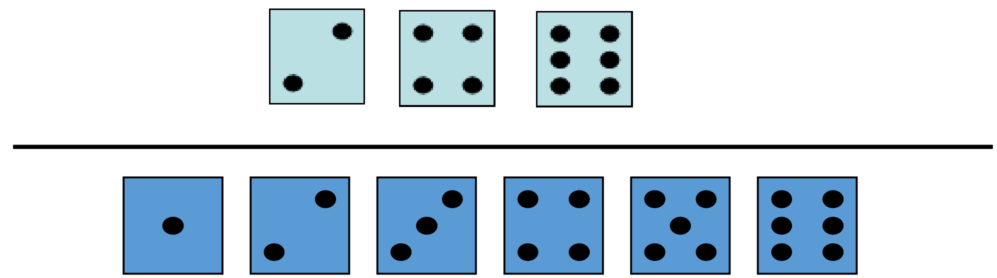
alle

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!

Ereignis *A*

Zufallsexperiment $\longrightarrow$

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

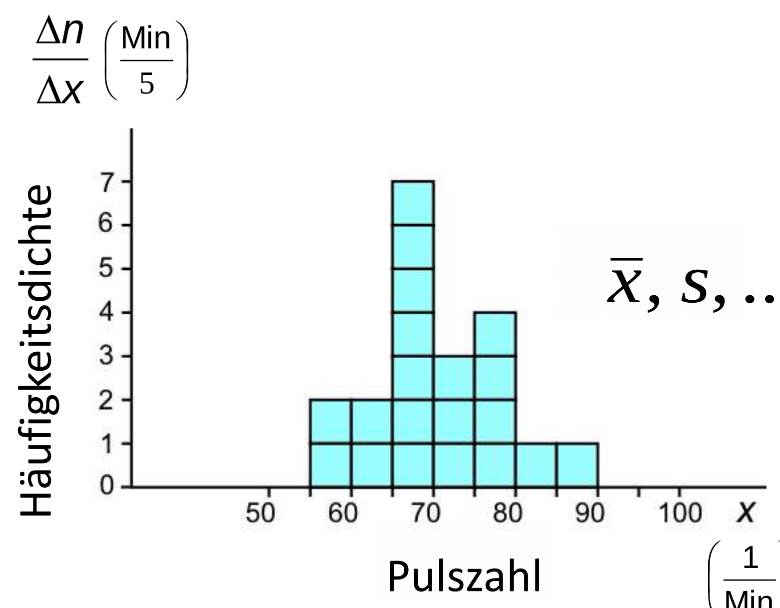
Population

Wahrscheinlichkeitsverteilung

?

?, ?, ...

Stichprobe



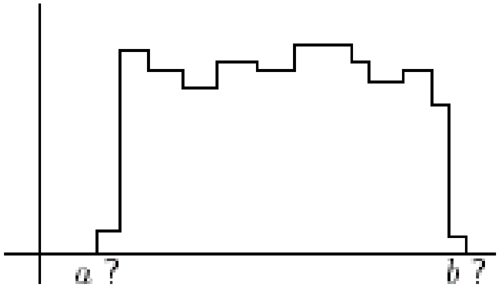
$\bar{x}, s, \ldots$

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
- ...

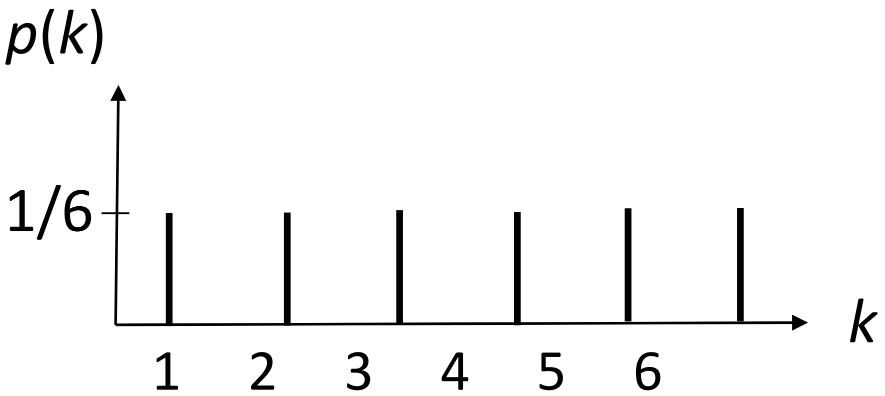
kontinuierliche Zufallsgröße

zB: Blutdruck, Körperhöhe,...

Diskrete Gleichverteilung

Beispiel:

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



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

weitere Beispiele:

Münzenversuch

Würfelexperiment mit einem Ikosaeder

Lageparameter der Verteilung

Es sei *X* eine diskrete Zufallsgröße mit Werten x_1,x_2… 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

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

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

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

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

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 $\rightarrow$ theoretische
Streuung Streuung

(Standardabweichung)

Normalverteilung

Verteilungsdichtefunktion:

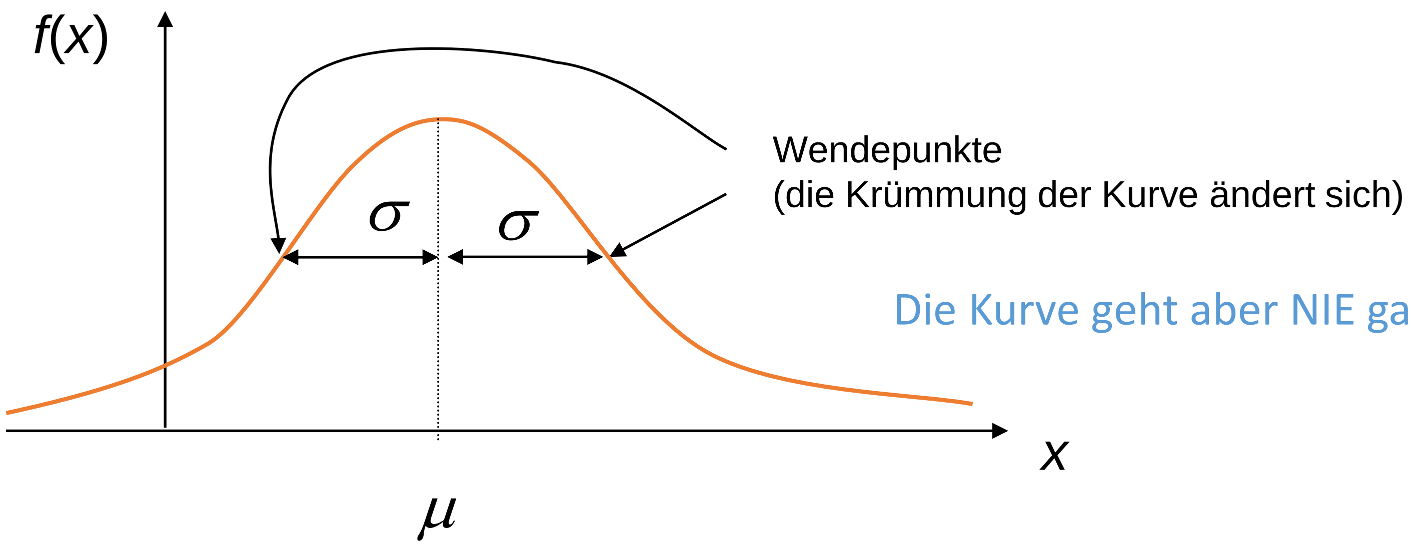
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Parameter der Normalverteilung:

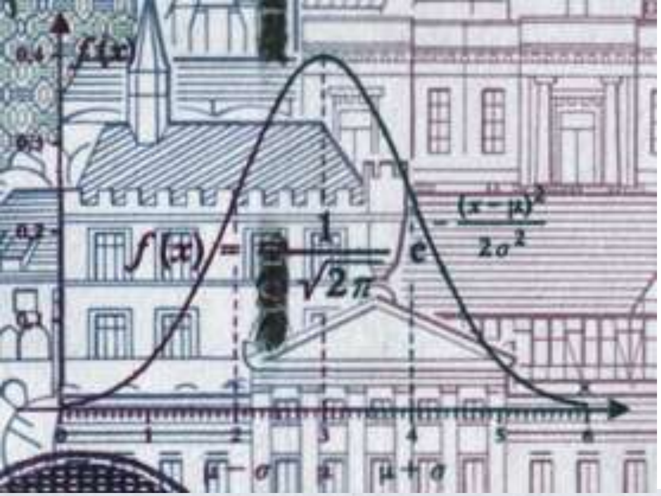
Erwartungswert: μ

Streuung: σ

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



Die Kurve geht aber NIE ganz zu Null!

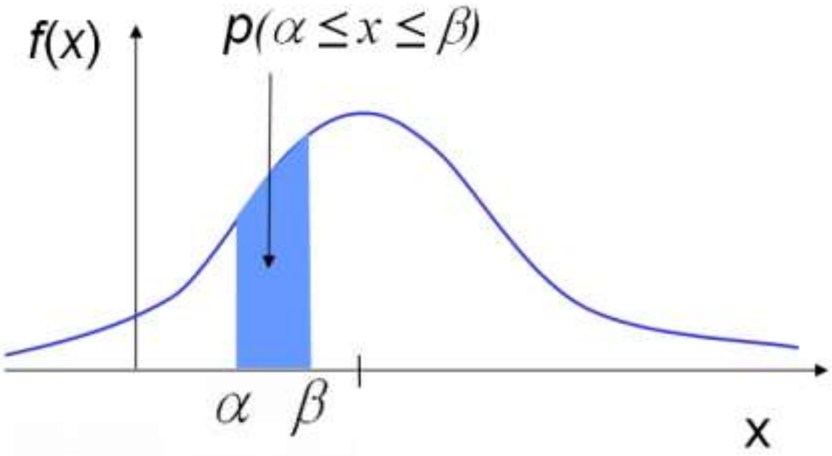


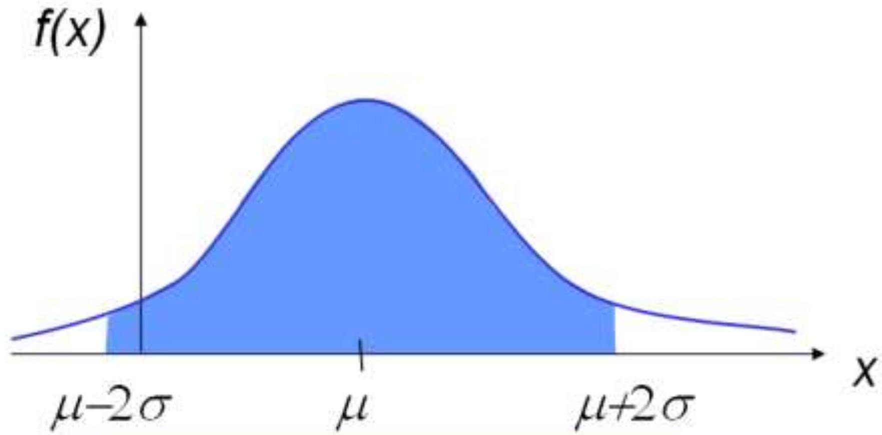
Normalverteilung (Gauss-Verteilung)

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

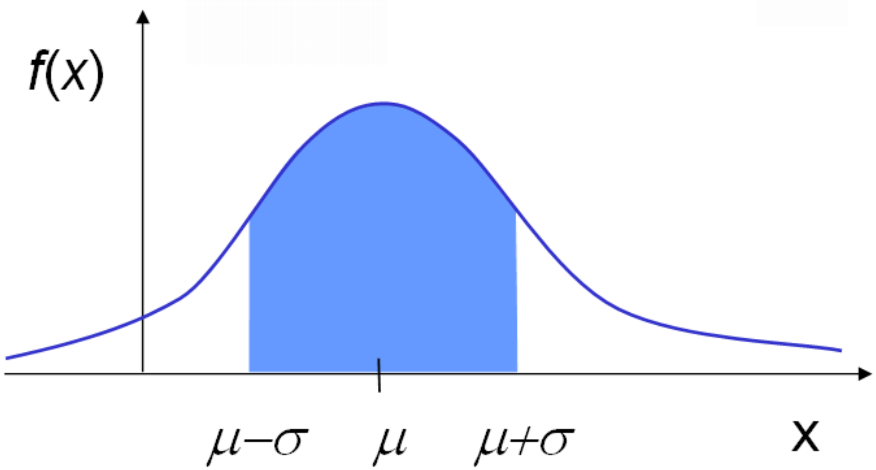
Normalverteilung

Wahrscheinlichkeit ist eine Oberfläche unter der Dichtefunktion!

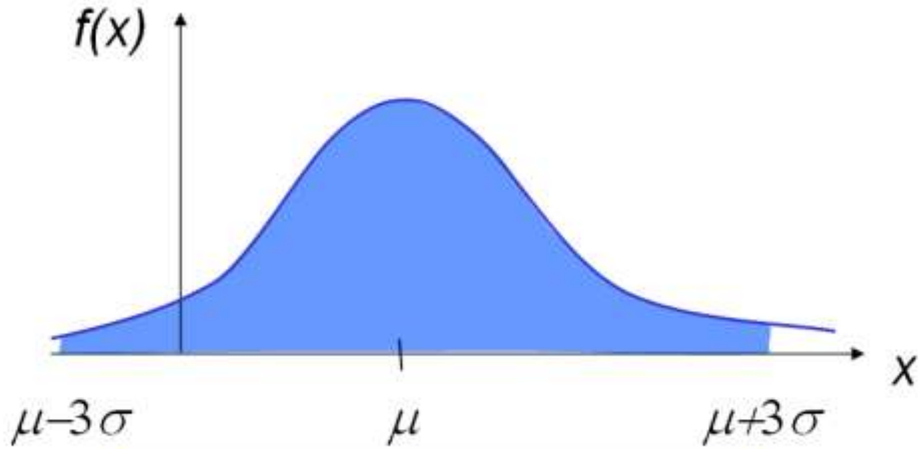




$p(\mu-2\sigma \leq x \leq \mu+2\sigma)=95\%$



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

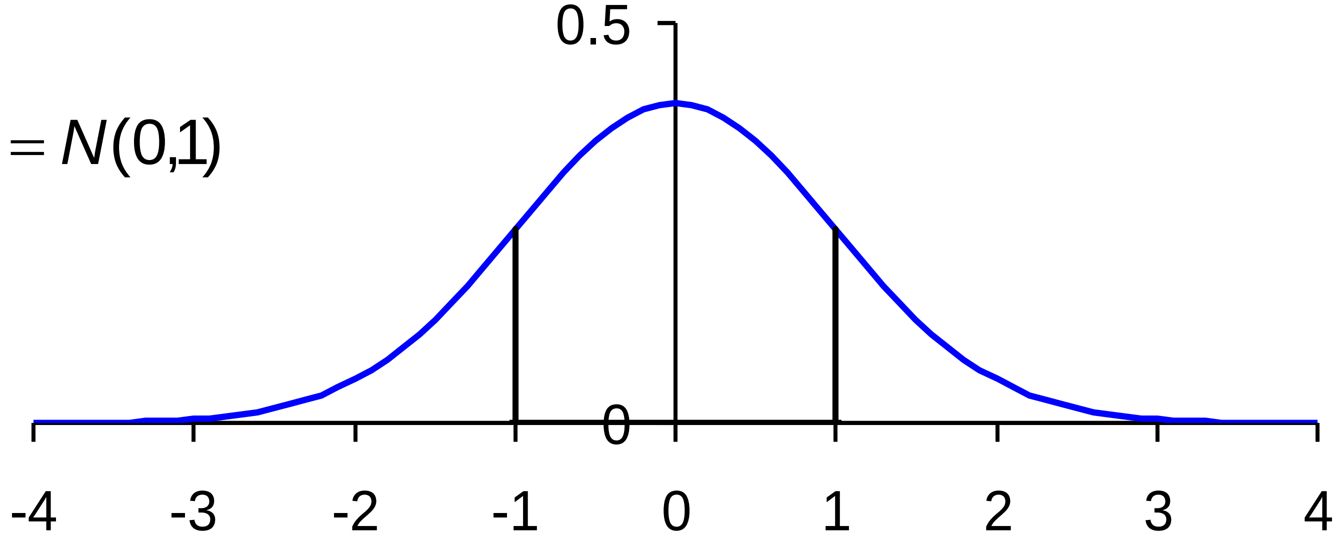


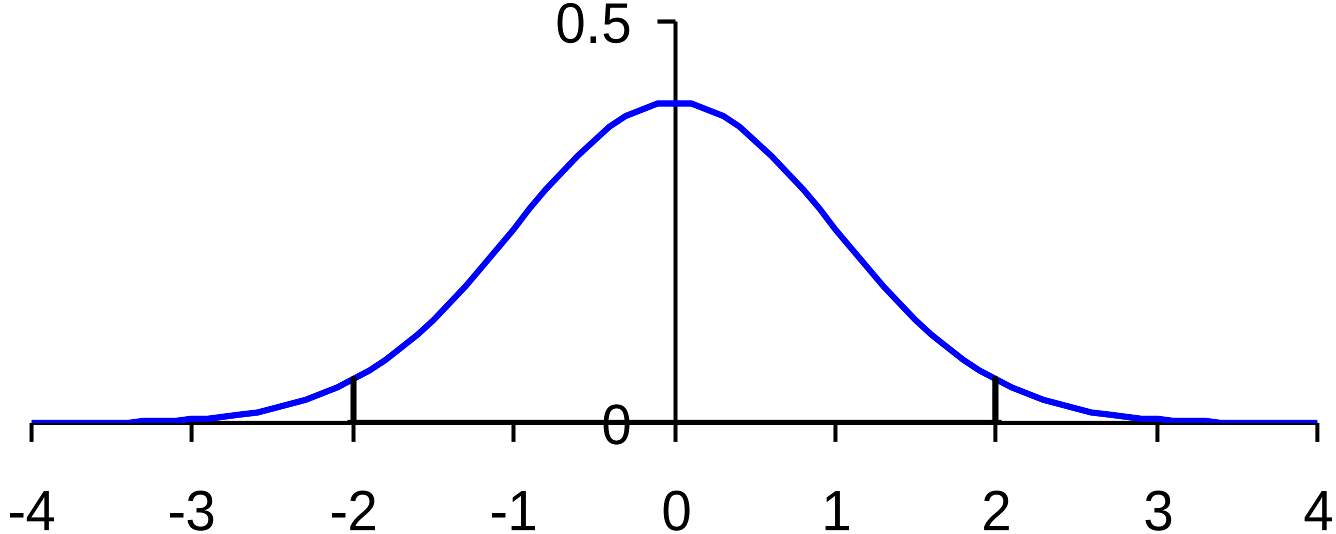
$p(\mu-3\sigma \leq x \leq \mu+3\sigma)=99,8\%$

Standard - Normalverteilung

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

$$\boxed{\begin{array}{l} \mu = 0 \\ \sigma = 1 \end{array}}$$



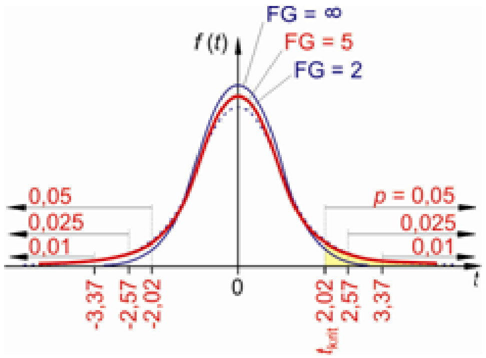


1. STATISTISCHE TABELLEN

t-VERTEILUNG

t-Verteilungs**familie**

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



„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 $n \rightarrow \infty$. $S_n = \sum_{i=1}^{n} x_i$

•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 = „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

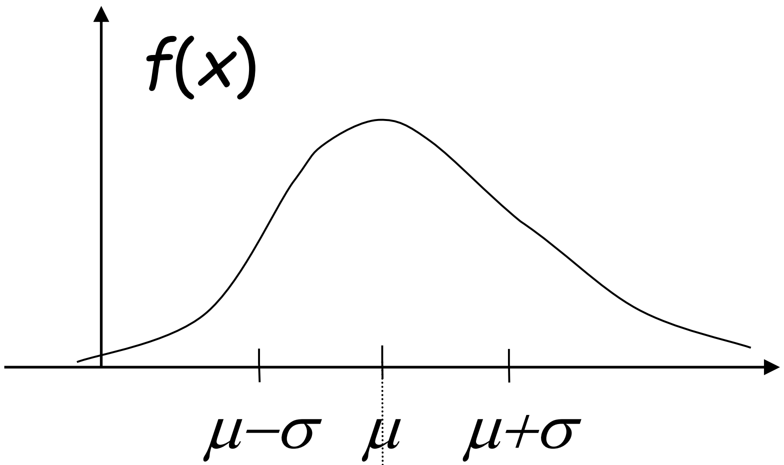
- Der Parameter wird mit einem Wert geschätzt.
- Relative Häufigkeit
  ist ein Schätzwert für die Wahrscheinlichkeit
- Durchschnitt
  ist ein Schätzwert für den Erwartungswert
- Standardabweichung
  ist ein Schätzwert für die Streuung
- Punktschätzungen sagen
  nichts über die Genauigkeit bzw. Sicherheit der Schätzung!

Intervallschätzungen

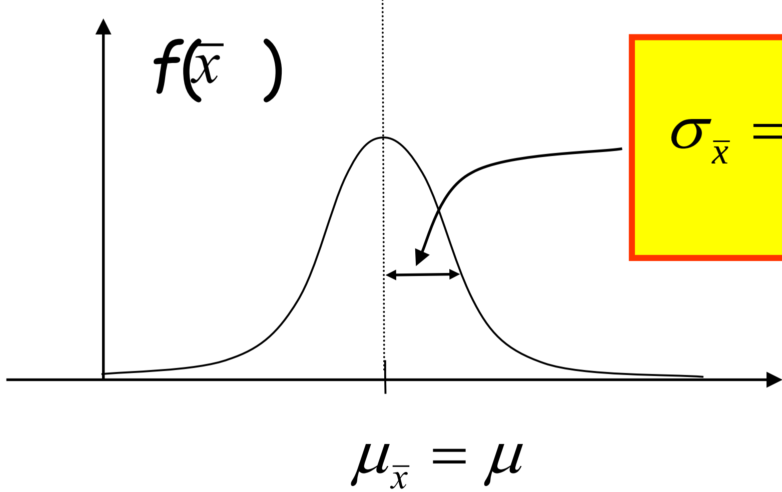
- Intervallschätzung oder Konfidenzschätzung gibt zu einer vorgewählten Sicherheitswahrscheinlichkeit γ, (Konfidenzniveau) ein Intervall (c_1,c_2) an, in dem der unbekannte Parameter (zB. μ oder σ) mit einer Wahrscheinlichkeit von mindestens γ liegt.

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

Konfidenzintervall für den Erwartungswert



x zB:Körperhöhe

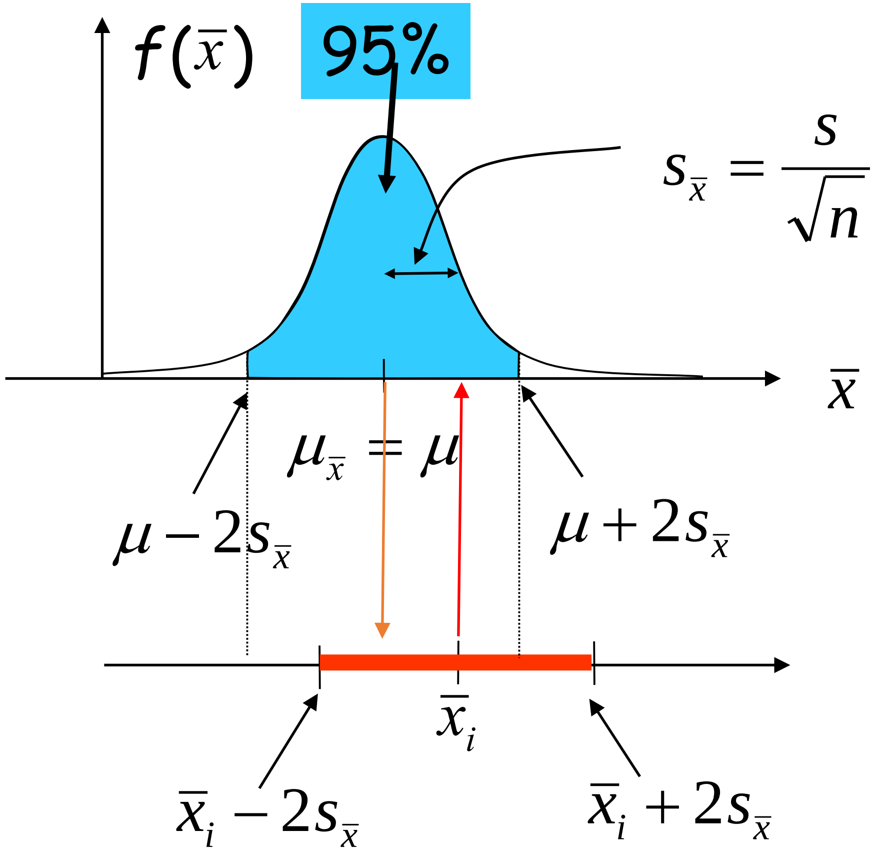


Standardfehler

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

$\bar{x}$ zB: durchschnittliche Körper-höhe 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

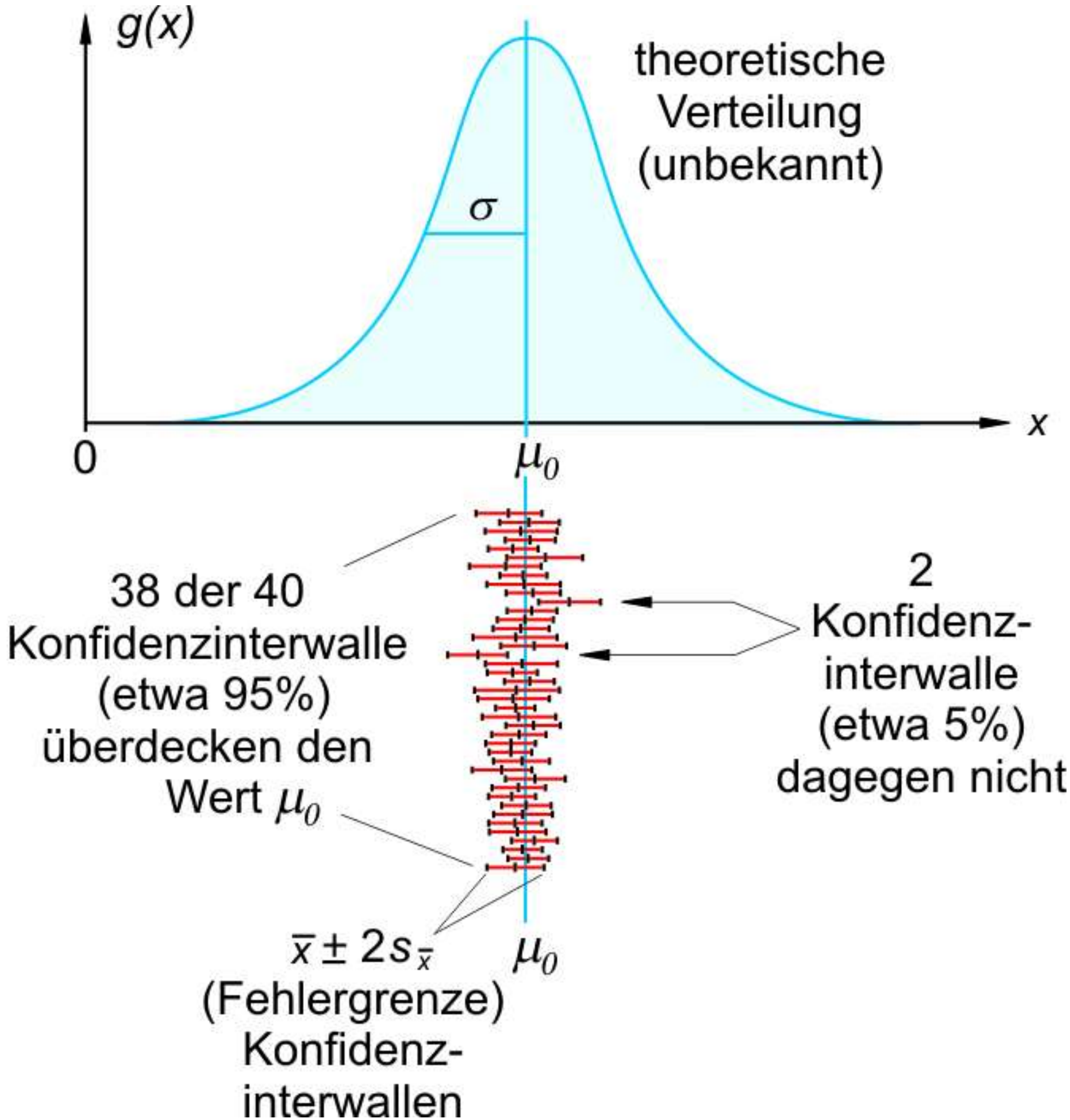
Eine ähnliche Ableitung gibt: μ liegt
-mit 68% Wahrscheinlichkeit im Intervall: $\bar{x} - s_{\bar{x}},\ \bar{x} + s_{\bar{x}}$

- 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$\to\infty$ dann $s_{\bar{x}} \to 0$



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:

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