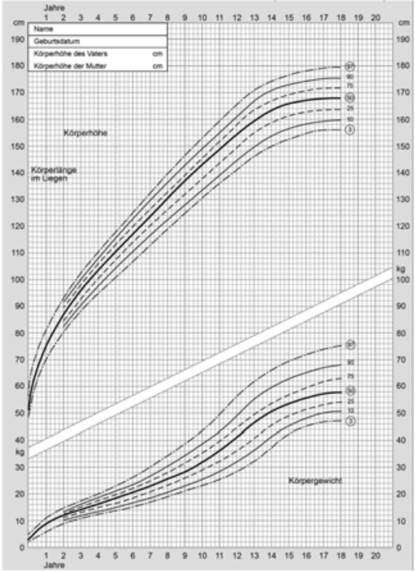
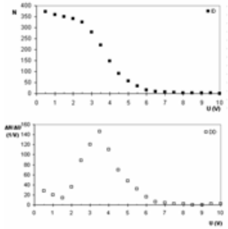


Deskriptive Statistik



KAD 2011.09.22

Element

Grundgesamtheit (Population):

Gesamtheit der Individuen (Elemente), deren Eigenschaften bei der Studie untersucht werden sollen. Die gesamte Menge der interessierenden Daten.

N = „unendlich"

Stichprobe:

Der für die Studie ausgewählte Teil der Population.

n = endlich

N >>*n* (Umfang)

über die man etwas aussagen möchte

die Stichproben-elemente sollen zufällig ausgewählt werden

Grundgesamtheit

Stichprobe

Zufall !

Unsicherheit !

deskriptive Statistik

induktive Statistik (schließende St. analytische St.)

Die deskriptive Statistik ist die Vorstufe zur induktiven Statistik

---

Wie hoch ist die normale Pulsfrequenz (einer Population)?

Merkmal: Pulsfrequenz

zufällige Erhebung einiger Elementen der Population: **Stichprobe**

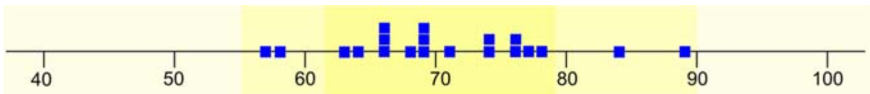
Daten der Stichprobe liegen in Form einer Urliste vor:

66, 56, 89, 63, 66, 69, 71, 68, 58, 69, 78, 66, 64, 84, 74, 76, 69, 77, 74, 76 (Einheit: 1/Min), oder:

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

„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

Verfeinern wir die Klassen noch weiter!

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

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:	$n = 20$

in Excel:

=frequency(...)
=Häufigkeit(...)

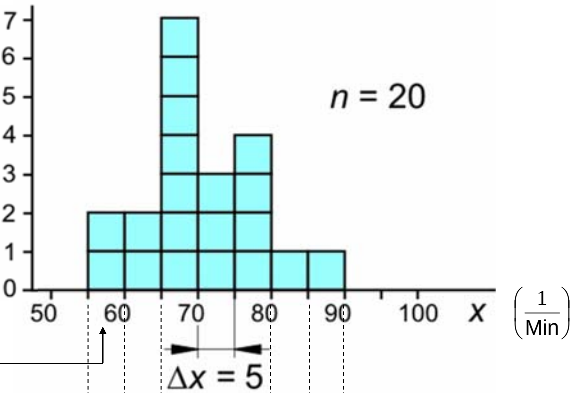
Die Grenzwerte und die Breiten der Klassen sind willkürlich.
Stellen wir diese Treppenfunktion dar!

Pr.Buch Tabelle 5

Häufigkeitsdichte

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

$$\left(\frac{1}{5\frac{1}{\text{Min}}}\right) = \left(\frac{\text{Min}}{5}\right)$$



Die Fläche unter der Treppenfunktion zwischen 55 und 60:

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

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

Anzahl der Messdaten in der Stichprobe

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:	$n = 20$

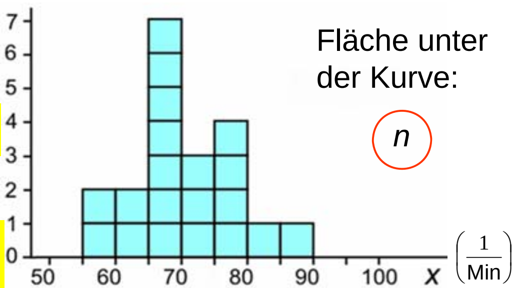
$$\frac{\Delta n}{\Delta x}\left(\frac{\text{Min}}{5}\right)$$

absolute

Häufigkeitsdichte-verteilung

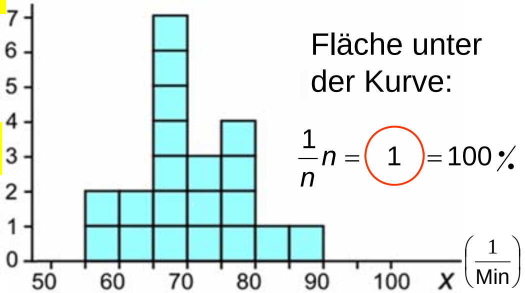
$$\frac{1}{n}\frac{\Delta n}{\Delta x}\left(\frac{1}{20}\frac{\text{Min}}{5}\right)$$

relative



Fläche unter der Kurve:

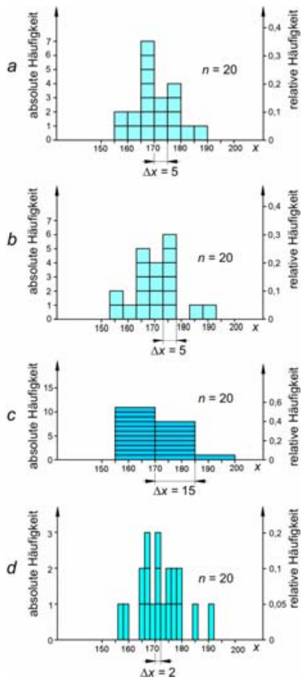
n



Fläche unter der Kurve:

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

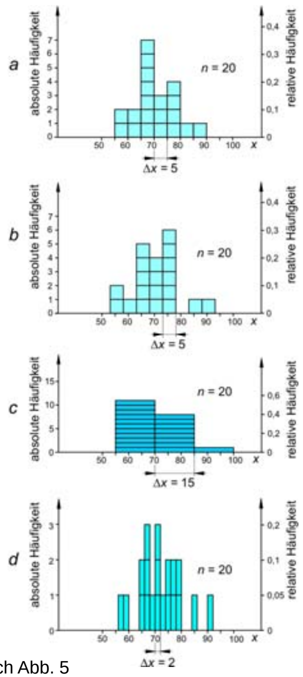
absolute Häufigkeits-dichte
(Histogramm)



relative Häufigkeits-dichte
(Histogramm)

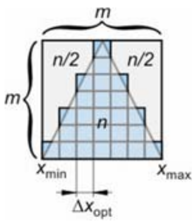
„Jedes Rechteck entspricht einem Messwert. „

Pr.Buch Abb. 5



Pr.Buch Abb. 5

Bestimmung der optimalen Klasseneinteilung



optimale Klassenanzahl m:

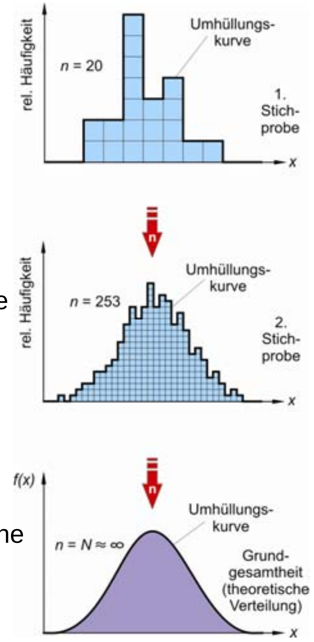
$$m^2 = 2n$$

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

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

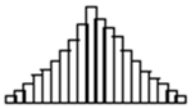
Pr.Buch Abb. 6

n vergrößert sich,
die Klassenbreite Δx kann verkleinert werden

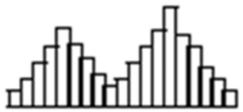
Bei großen Stichproben ergibt die empirische Verteilungsfunktion **eine sehr gute Näherung** der theoretischen Verteilungsfunktion. (Die Stichprobe ist „gleich" der Grundgesamtheit.)

Analyse von Häufigkeitsverteilungen

homogene symmetrische Stichprobe:

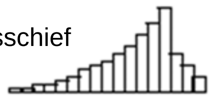


heterogene Stichprobe:

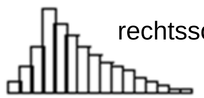


homogene nichtsymmetrische Stichproben:

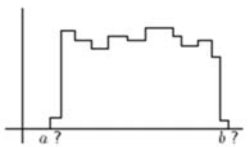
linksschief



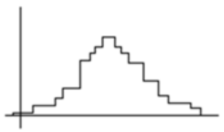
rechtsschief



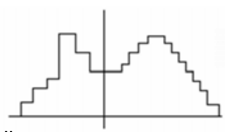
Vermutung:



Gleichverteilung?



Normalverteilung?



Überlagerung von zwei Normalverteilungen?

Lageparameter. Charakterisierung das Zentrum der Daten

Durchschnittswert (der arithmetische Mittelwert)

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

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

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

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

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

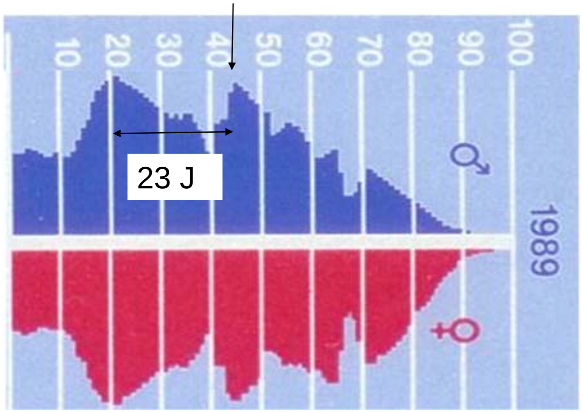
$$x_{\text{med}} = \begin{cases} x_{(n+1)/2} & \text{falls } n \text{ ungerade} \\ (x_{n/2} + x_{(n/2+1)})/2 & \text{falls } n \text{ gerade} \end{cases}$$

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

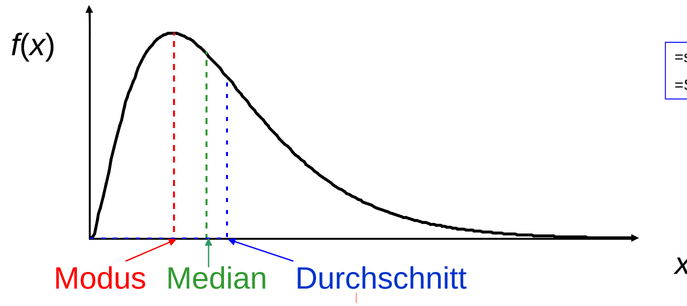
Altersaufbau der deutschen Bevölkerung

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

1989-43=1946



Linkssteile bzw. <u>rechtschiefe</u> Verteilung



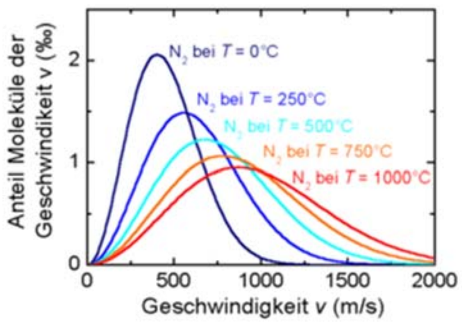
=skew(...) > 0
=Schiefe(...) > 0

z.B. Einkommensverteilungen in einem Land:

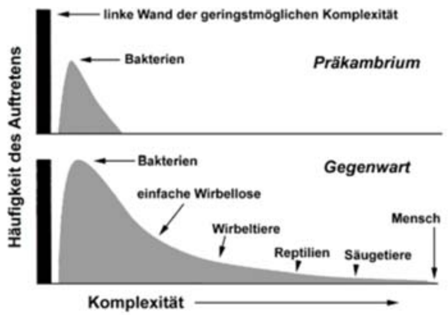
Der Großteil der Bevölkerung verdient relativ wenig, während es nur wenig Leute gibt, die sehr viel verdienen.

Weitere Beispiele

Maxwell-Boltzmann-Verteilung

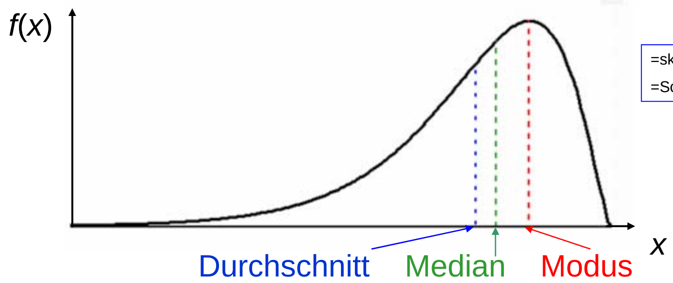


Komplexität der Tiere



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

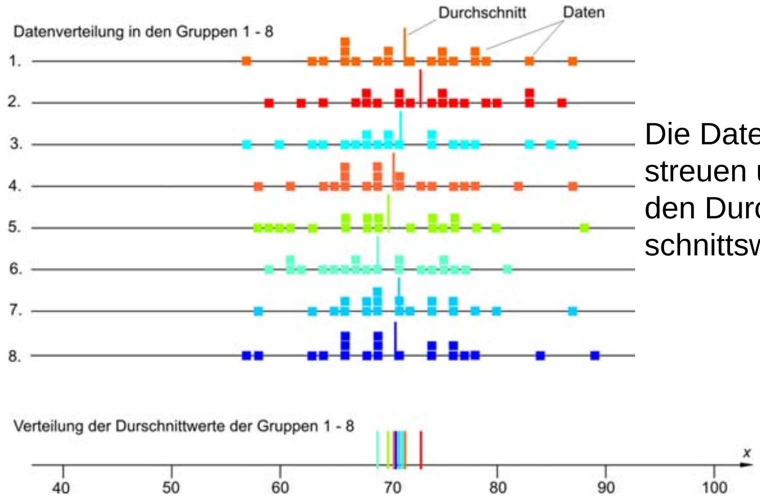
<u>Linksschiefe</u> bzw. rechtssteile Verteilung



=skew(...) < 0
=Schiefe(...) < 0

z.B. Dauer einer Schwangerschaft

Daten und ihre Durchschnittswerte



Die Daten streuen um den Durch-schnittswert.

Pulsfrequenzen (1/Min)

Pr.Buch Abb. 10

Streuungsmasse (Variabilitätsmaße, Variationsmaße)
Mass für die Streubreite von Daten

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):

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

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

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

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

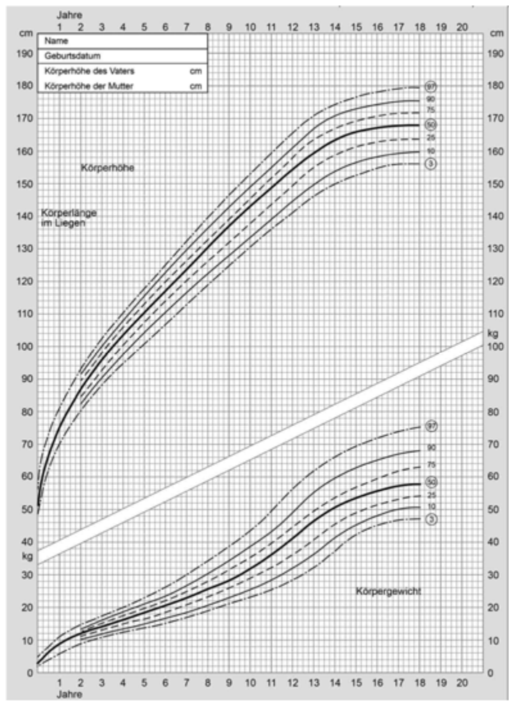
=Quartil(...)

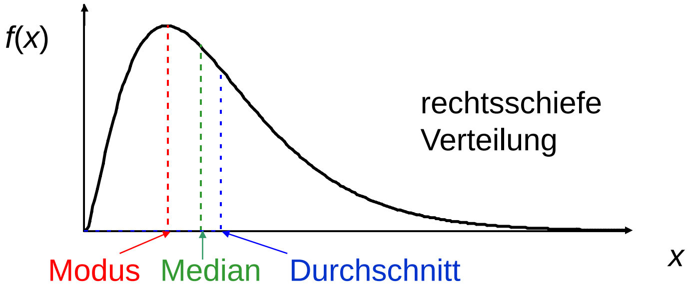
mit Wörter: z.B. **Dezile**
Durch Dezile (lat. „Zehntelwerte") wird die Verteilung in 10 gleich große Teile zerlegt. Unterhalb des dritten Dezils liegen 30 % der Verteilung.

Perzentilenkurven sind ein Werkzeug für den Arzt.

Wachstums- und Gewichtskurven für Mädchen

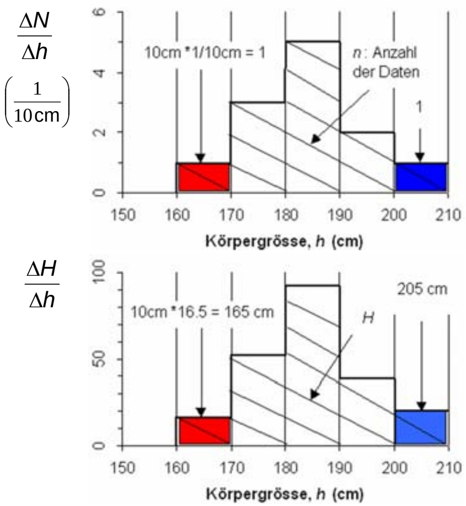
=percentile(...)
=Quantil(...)





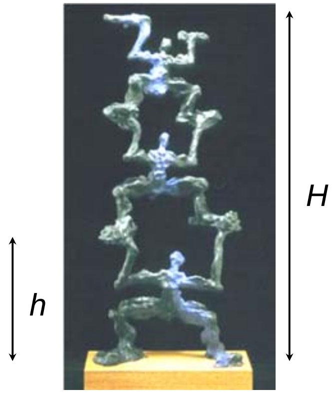
Skalentypen	zulässige Lage-Parameter	zulässige Streuungs-Parameter
Nominalskala	Modus	–
Ordinalskala	Modus, Median	–
numerische Skalen	Modus, Median, Durchschnittswert	Spannweite, Quartilabstand, Standardabweichung

Häufigkeitsverteilung



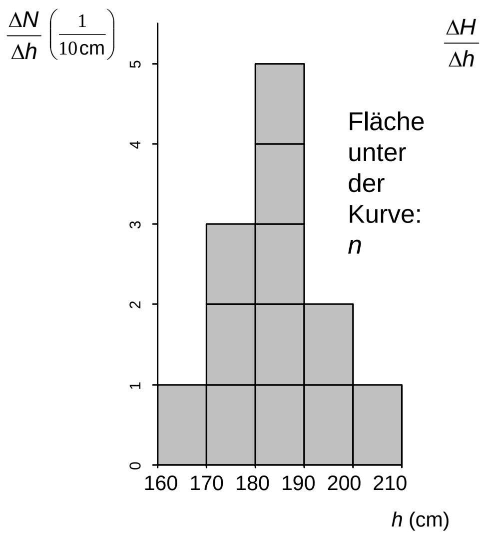
h: Körperhöhe

H: kollektive Höhe, Gesamthöhe



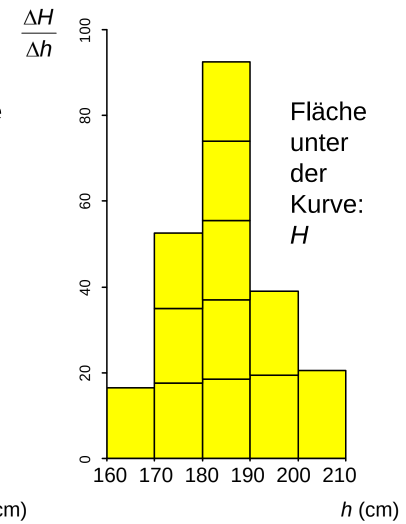
Spektrum als eine spezielle Häufigkeitsverteilung

Häufigkeitsdichte



Fläche unter der Kurve: *n*

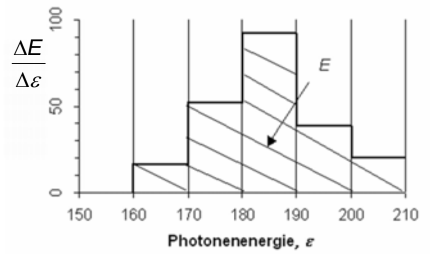
Spektrum



Fläche unter der Kurve: *H*

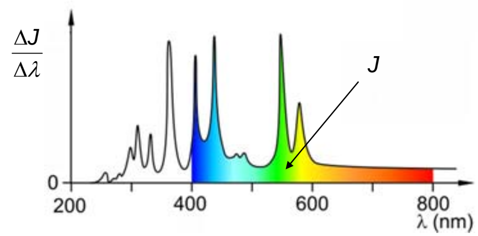
Emissionsspektrum:

wie verteilt sich die emittierte Energie über die Photonenenergien

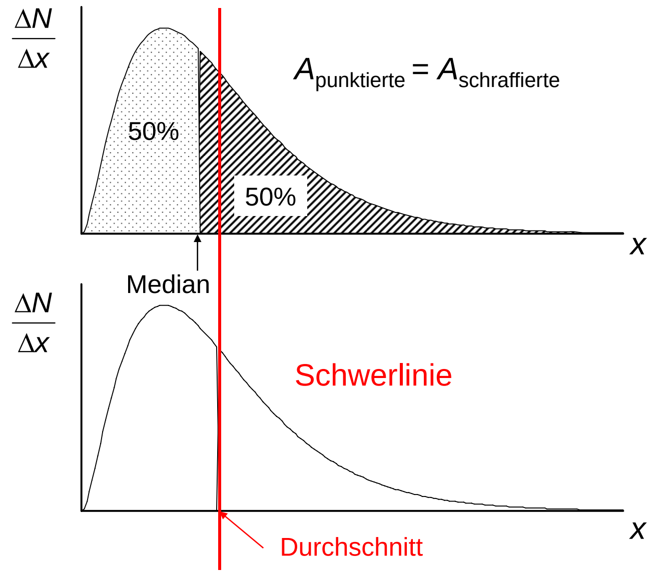


charakteristische Größe des Energietransports: **Intensität**

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

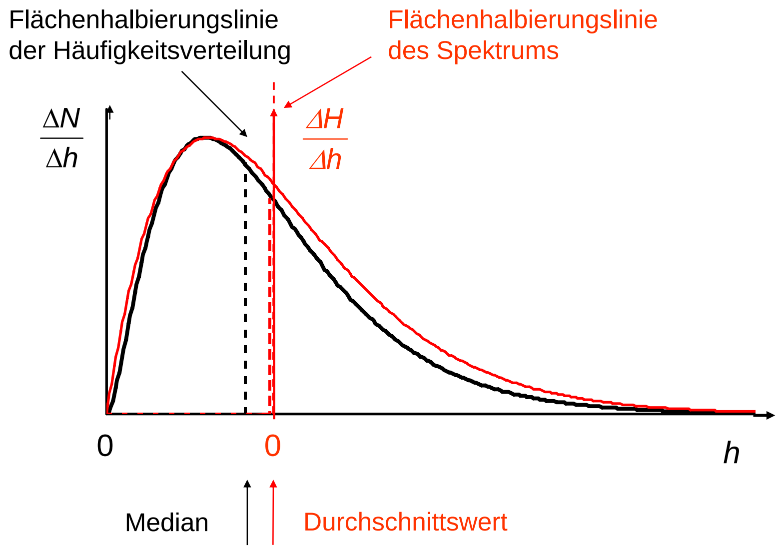


Position des Medians und des Durchschnitts einer Verteilung (1)



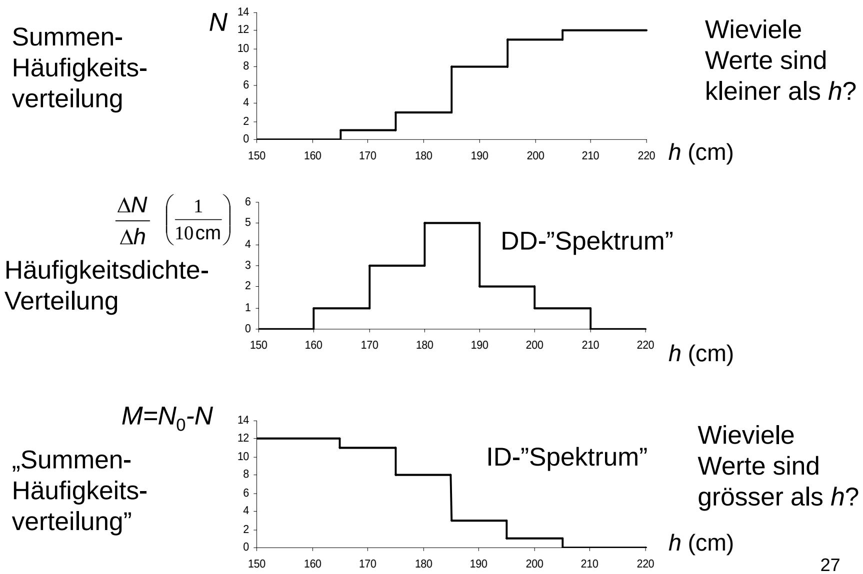
25

Position des Medians und des Durchschnitts einer Verteilung (2)



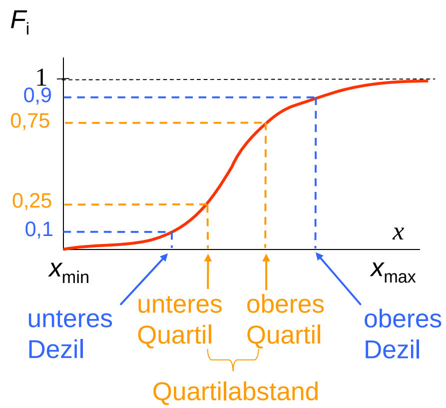
26

Summen- (kumulierte/kumulative) Häufigkeitsverteilung

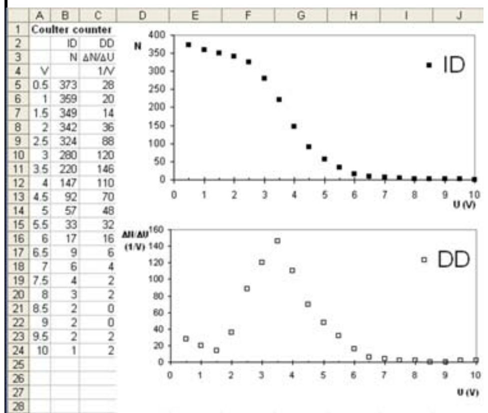


27

Quantile und die relative Summenhäufigkeits-verteilung



Coulter Zähler



	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			