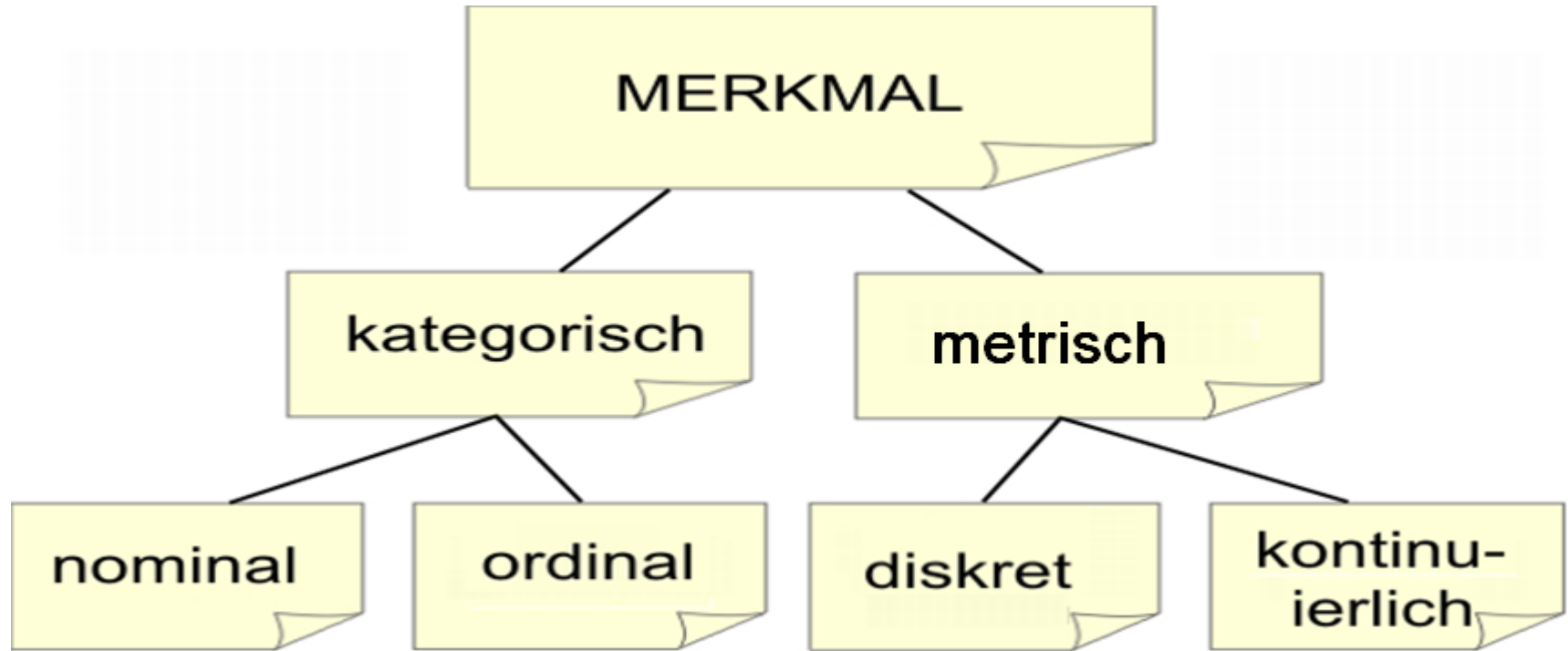


Widerholung: Klassifizierung der Merkmale



Übung mit Komparativ und Superlativ

gut



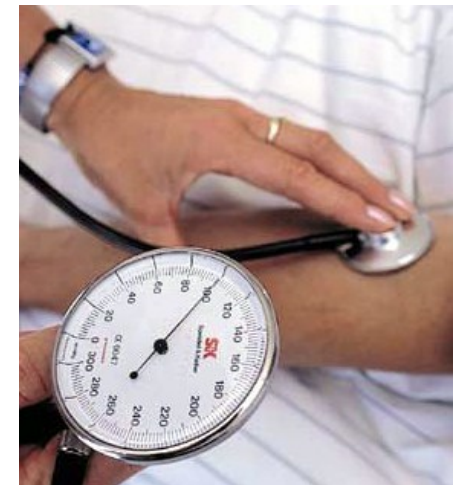
gut





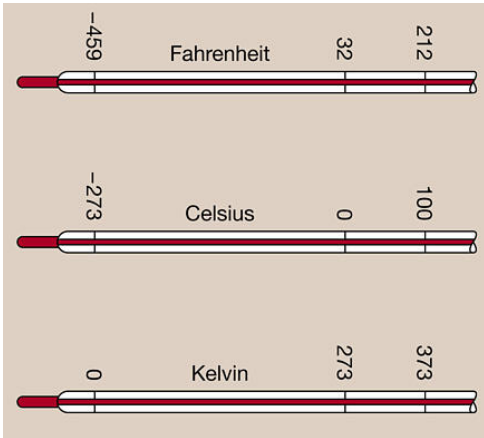
besser

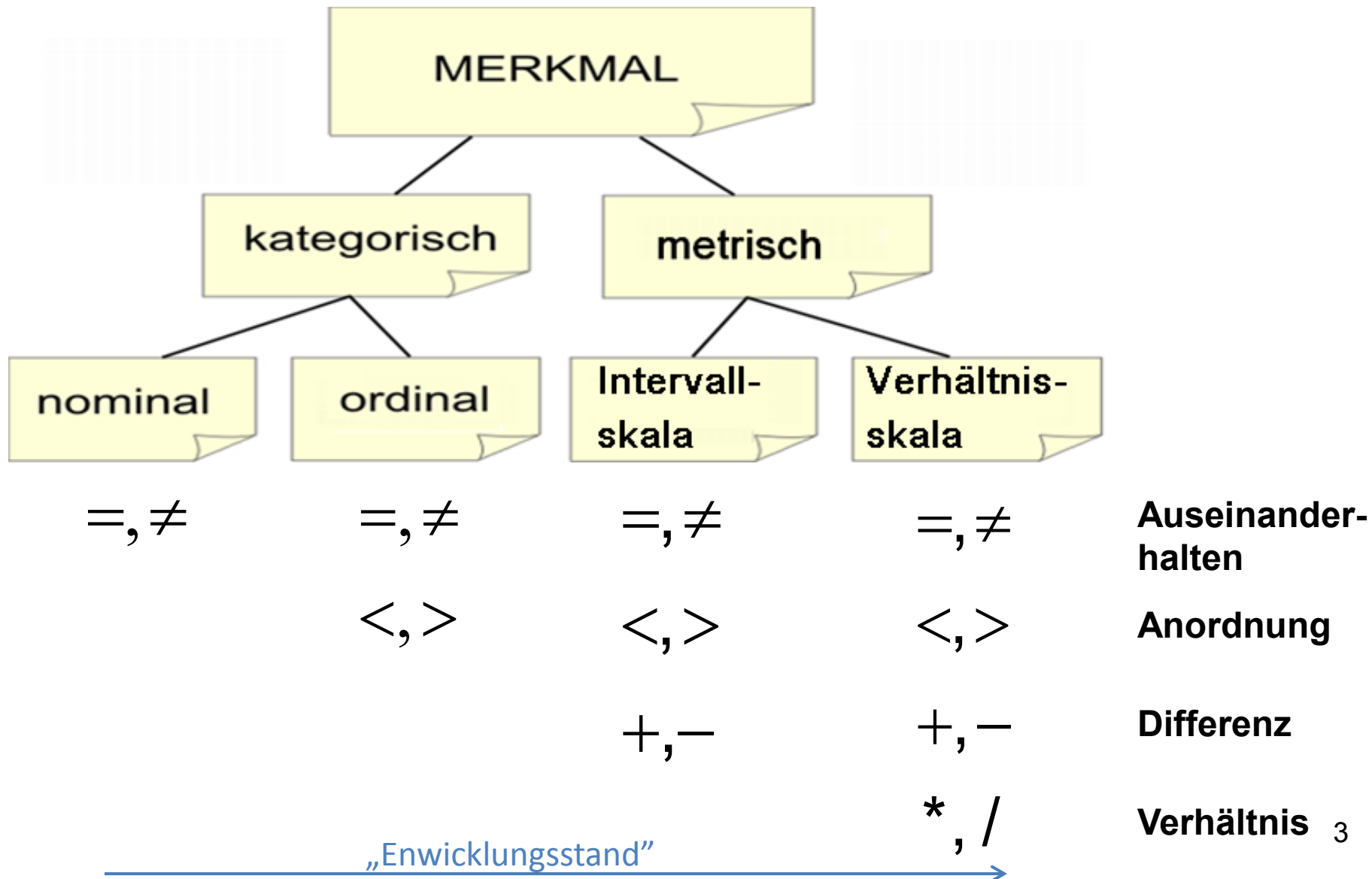


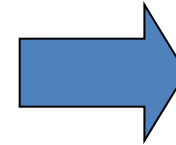
am besten



Skalentypen der metrischen Merkmale

	diskret	kontinuierlich																																																	
<div>Intervall- skala</div> <div>definierte Differenz, „kein“ 0 Punkt</div>	<div>Tage in einem Kalender</div> <table><tr><th colspan="7">Feb - 2009</th></tr><tr><th>Mo</th><th>Di</th><th>Mi</th><th>Do</th><th>Fr</th><th>Sa</th><th>So</th></tr><tr><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td><td>1</td></tr><tr><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr><tr><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td></tr><tr><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td></tr><tr><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>1</td></tr></table>	Feb - 2009							Mo	Di	Mi	Do	Fr	Sa	So	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	1	<div>Tempe- ratur in °C</div> 
Feb - 2009																																																			
Mo	Di	Mi	Do	Fr	Sa	So																																													
26	27	28	29	30	31	1																																													
2	3	4	5	6	7	8																																													
9	10	11	12	13	14	15																																													
16	17	18	19	20	21	22																																													
23	24	25	26	27	28	1																																													
<div>Verhältnis- skala</div> <div>definiertes Verhältnis, 0 Punkt</div>	<div>Anzahl der Zähne</div> 	<div>Tempe- ratur in K</div> 																																																	





Ein
Element

Stichprobe:

Grundgesamtheit (Population):

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

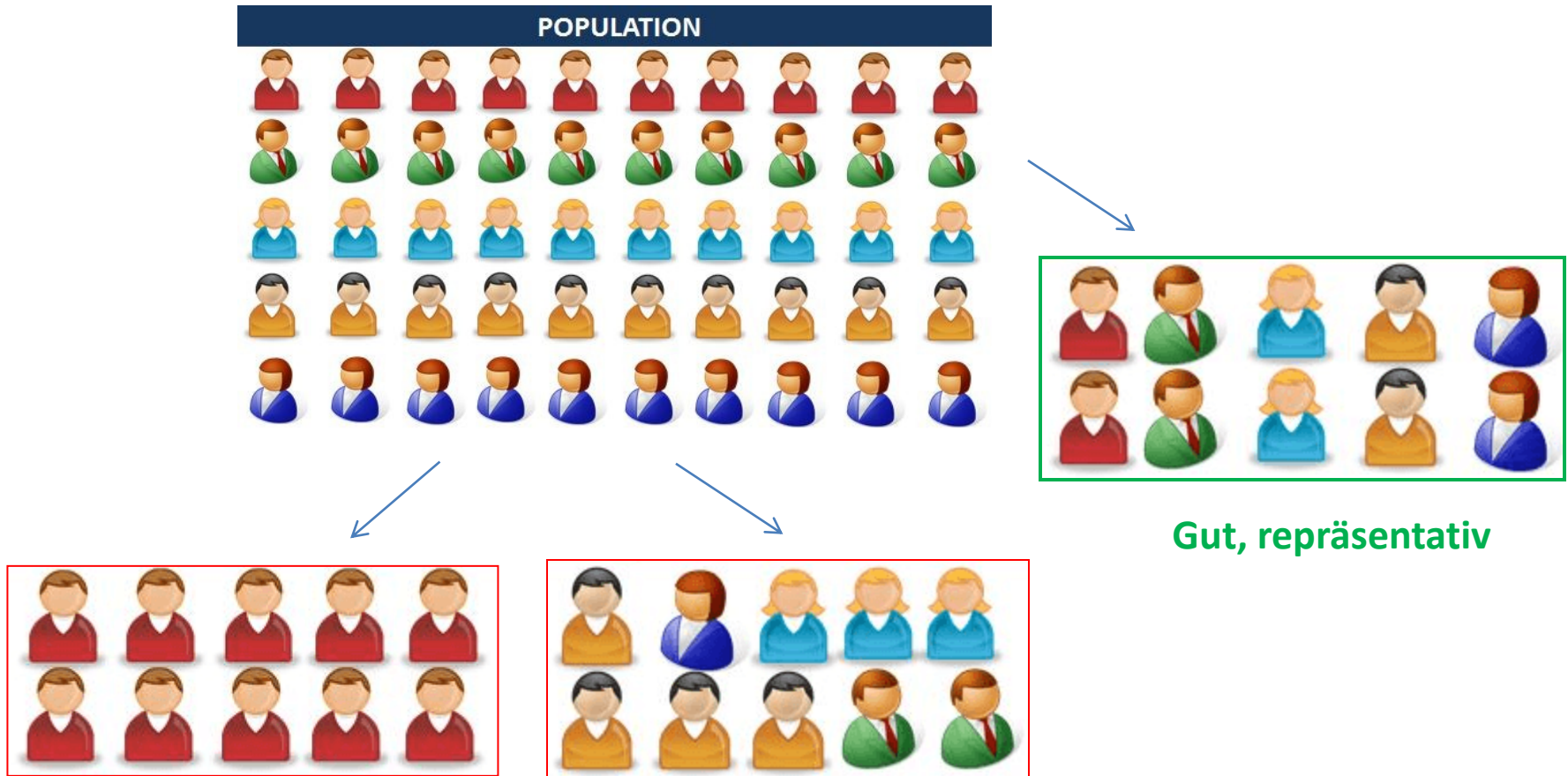
$N = \infty$ oder ungeheuer groß

Der für die Studie
ausgewählte Teil der
Population.

$n \ll N$

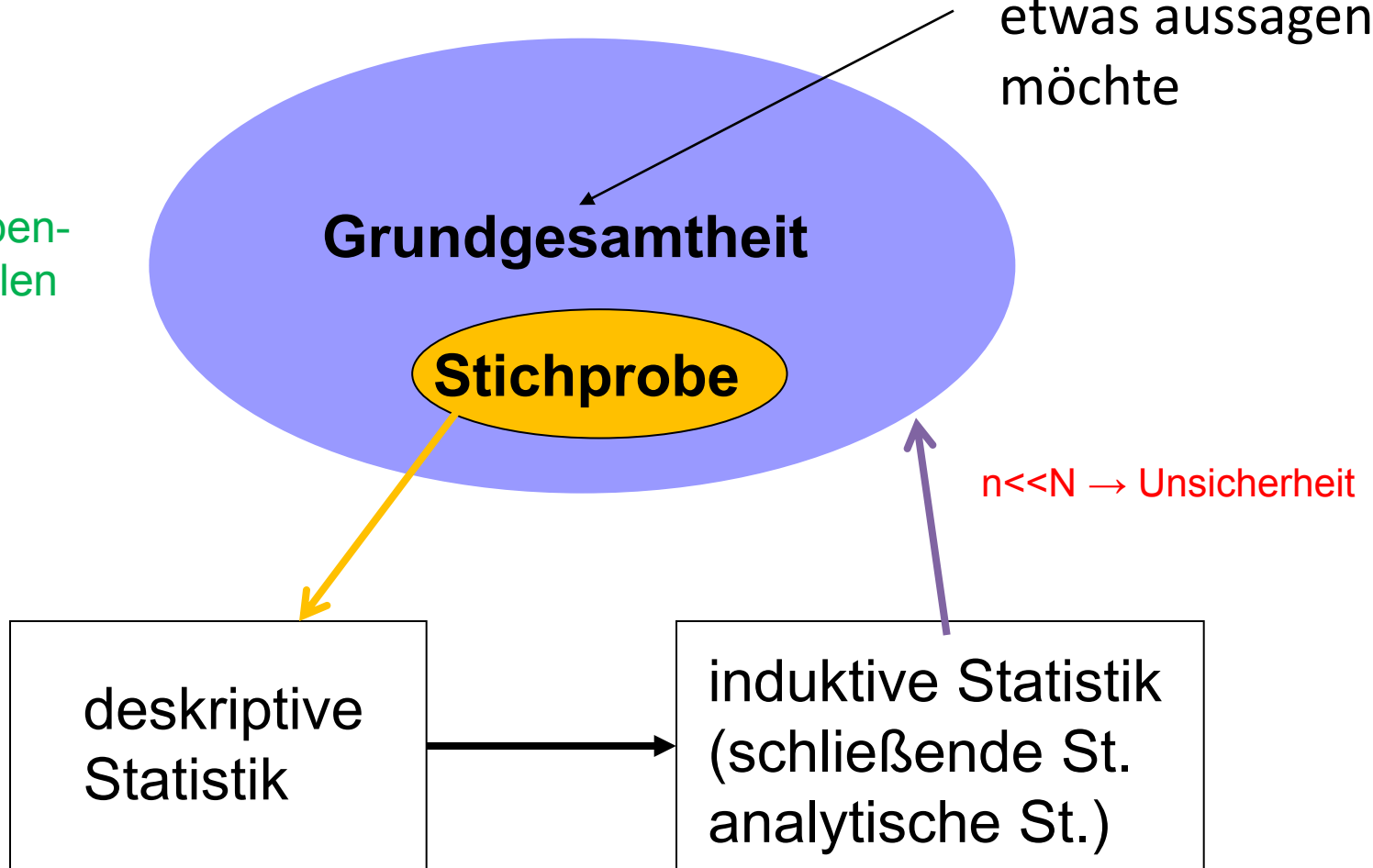
*Umfang d. Stichprobe =
Anzahl d. Daten*

Wir brauchen eine **repräsentative Stichprobe**



die Stichproben-
elemente sollen
zufällig
ausgewählt
werden

über die man
etwas aussagen
möchte



Frage: Wie hoch ist die normale Pulsfrequenz?

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



Stichprobe

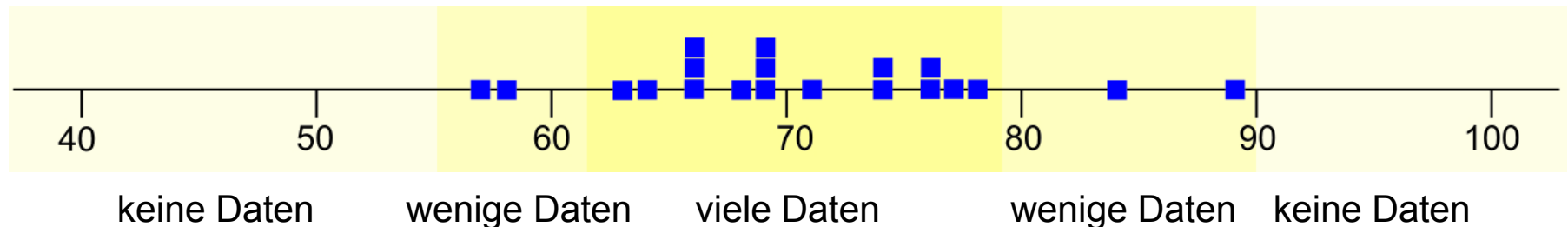
66	56	89	63	66	69	71	68	58	69
78	66	64	84	74	76	69	77	74	76

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!



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.

Häufigkeitsdichte

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

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

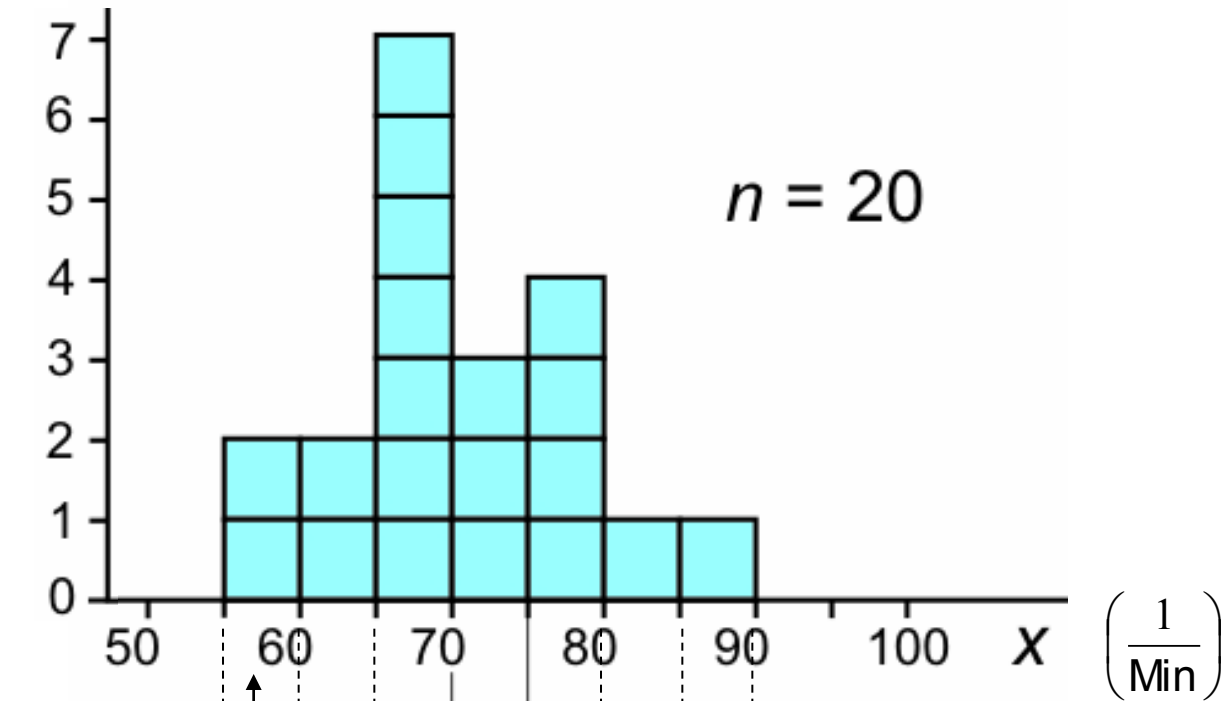
n.B. „Stück“ als Einheit lässt man oft weg.

Die Fläche unter der Treppenfunktion zwischen 55 und 60:

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

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

Anzahl der Messdaten in der Stichprobe

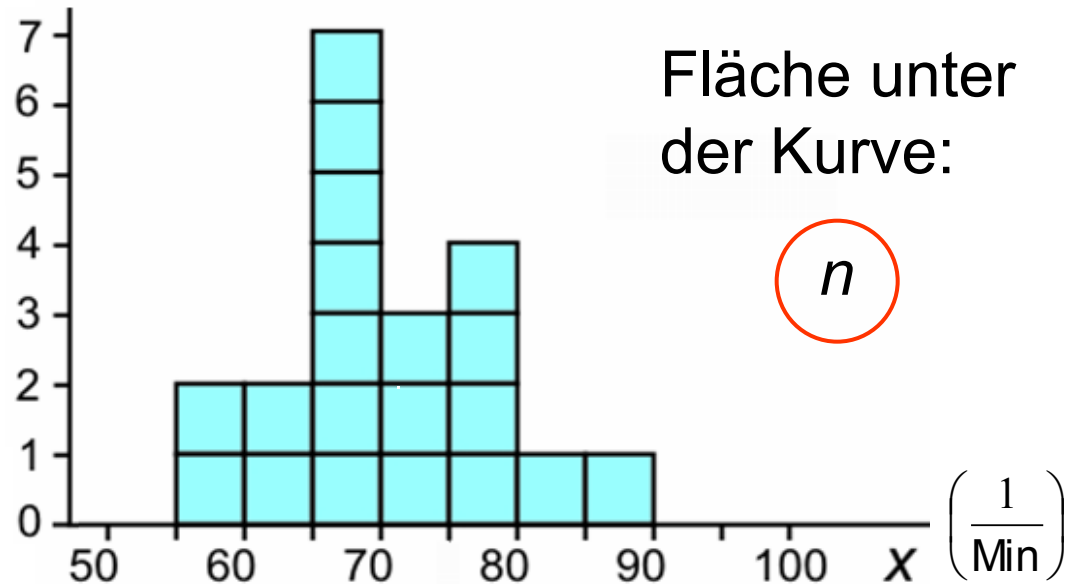


KLASSEN	GRENZEN	HÄUFIGKEIT
	$55 \leq x_i < 60$	2
	$60 \leq x_i < 65$	2
	$65 \leq x_i < 70$	7
	$70 \leq x_i < 75$	3
	$75 \leq x_i < 80$	4
	$80 \leq x_i < 85$	1
	$85 \leq x_i < 90$	1
	insgesamt:	$n = 20$

Häufigkeitsdichte- verteilung

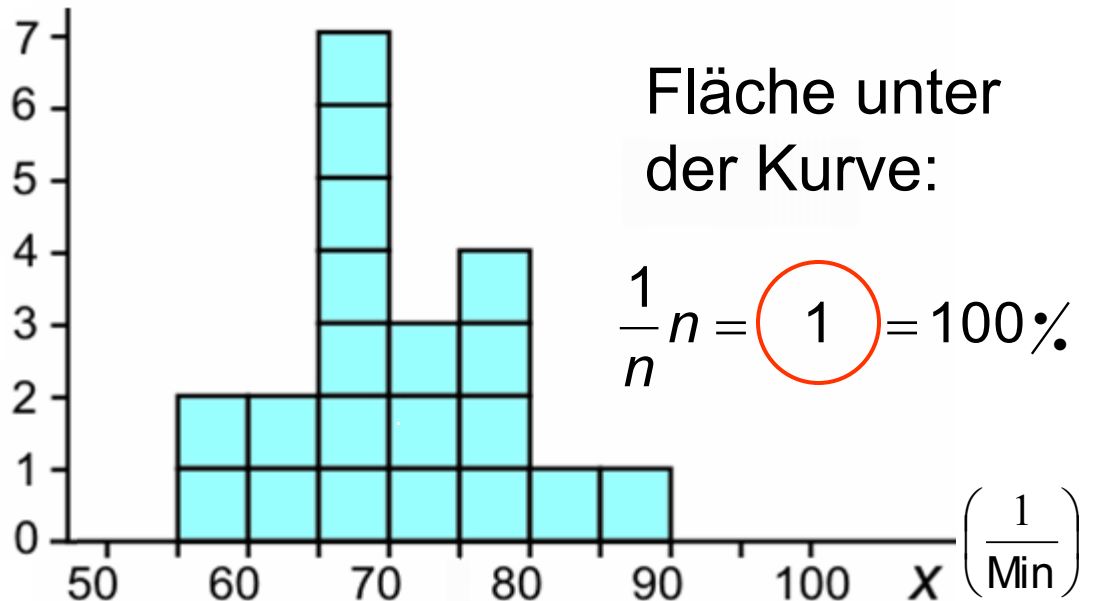
absolute

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

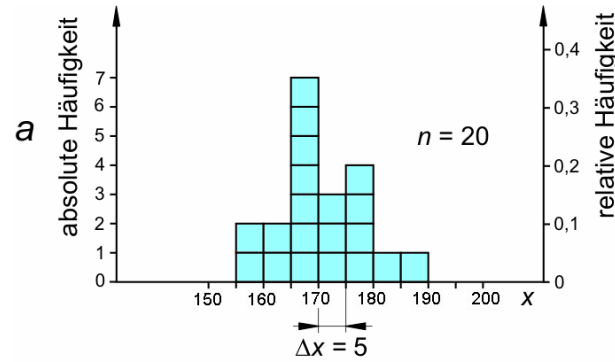


$$\frac{1}{n} \frac{\Delta n}{\Delta x} \left(\frac{1}{20} \frac{\text{Min}}{5} \right)$$

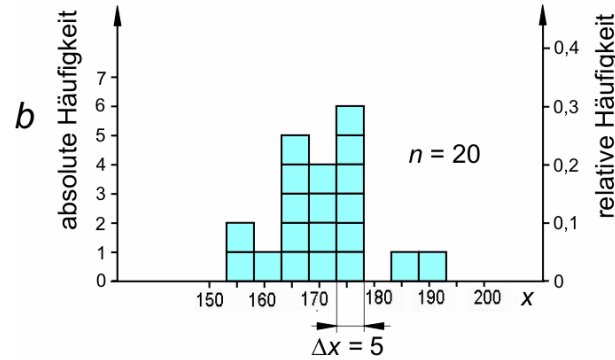
relative



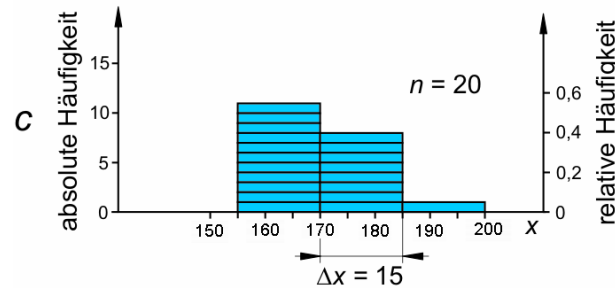
Die Klassenbreite kann das Aussehen des Histogramms wesentlich beeinflussen, wenn die Datenmenge nicht groß genug ist.
 In diesem Fall gibt es auch eine relativ hohe Instabilität des Histogramms



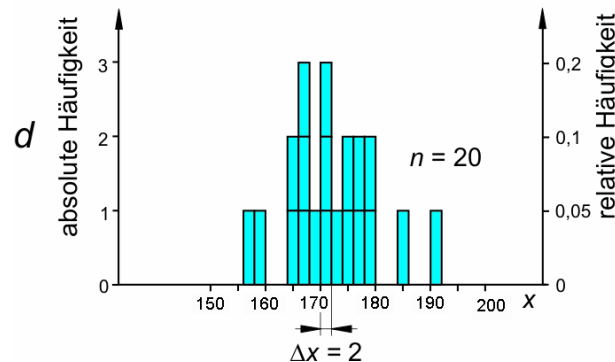
Selbe Grundgesamtheit, 2 Stchproben

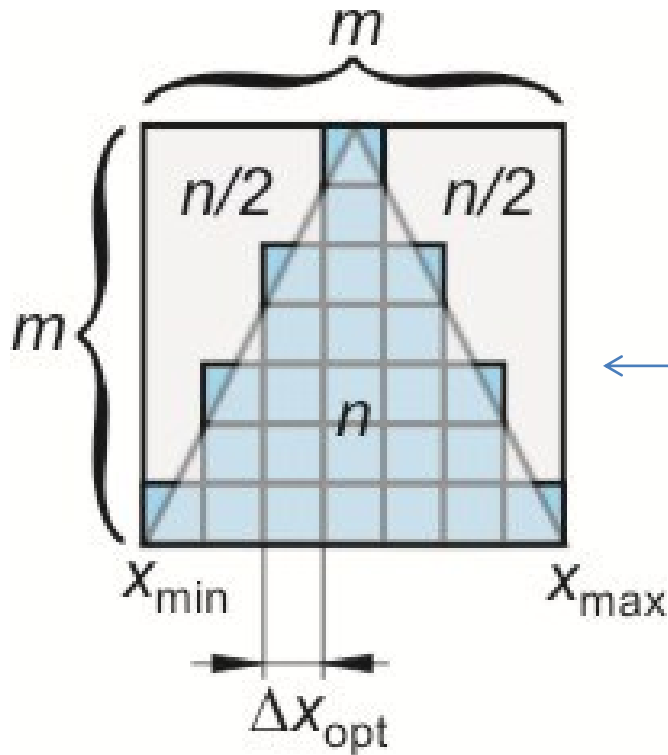


Zu große Klassenbreite



Zu kleine Klassenbreite





Bestimmung der optimalen Klasseneinteilung

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

optimale Klassenanzahl m Stück:

$$m^2 = 2n$$

$$m = \sqrt{2n}$$

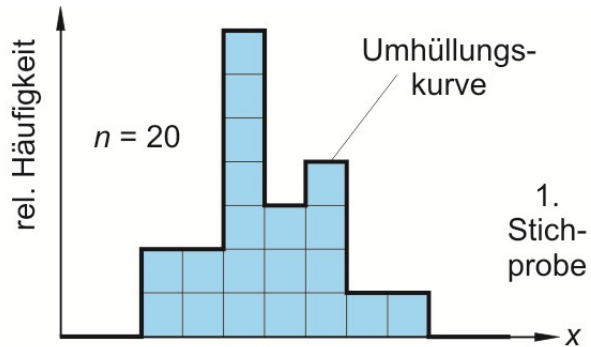
$$m = \sqrt{40} = 6.3$$

optimale Klassenbreite Δx :

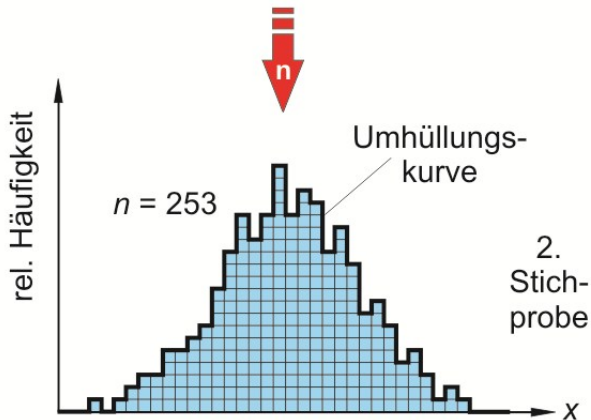
$$\Delta x = \frac{x_{\max} - x_{\min}}{m}$$

$$\Delta x = \frac{89 - 56}{6.3} = 5.2$$

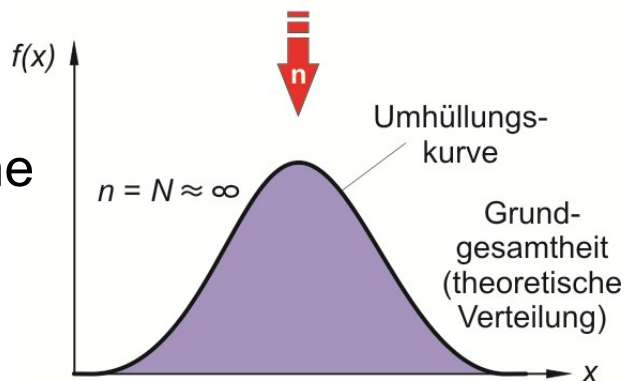
empirische
Funktion



empirische
Funktion



theoretische
Funktion

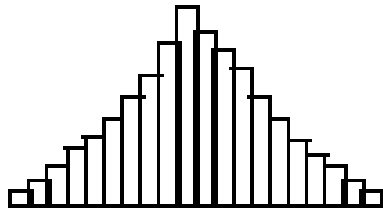


n vergrößert sich,
die Klassenbreite Δx kann
verkleinert werden

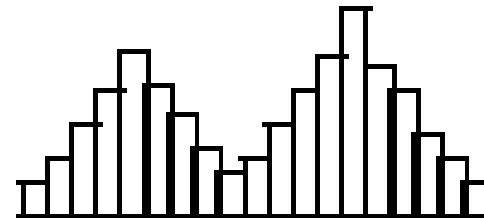
Bei großen Stichproben ergibt die empirische Verteilungsfunktion **eine sehr gute Näherung** der theoretischen Verteilungsfunktion. (Die Stichprobe ist „fast gleich“ der Grundgesamtheit.)

Analyse von Häufigkeitsverteilungen

homogene symmetrische Stichprobe:

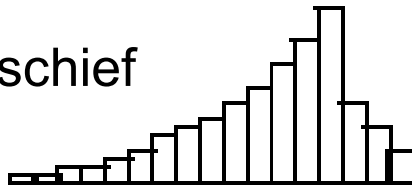


heterogene Stichprobe:

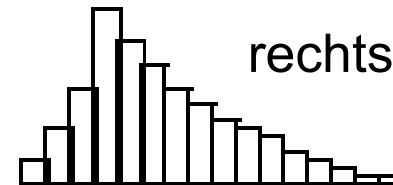


homogene nichtsymmetrische Stichproben:

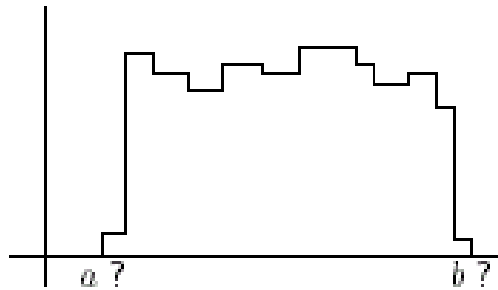
linksschief



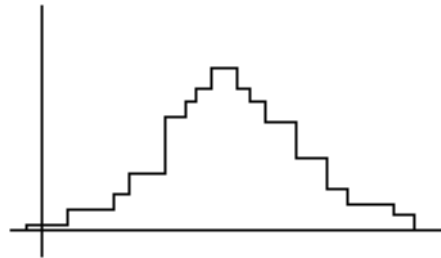
rechtsschief



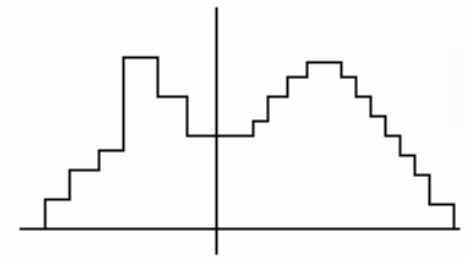
Vermutungen macht man auch:



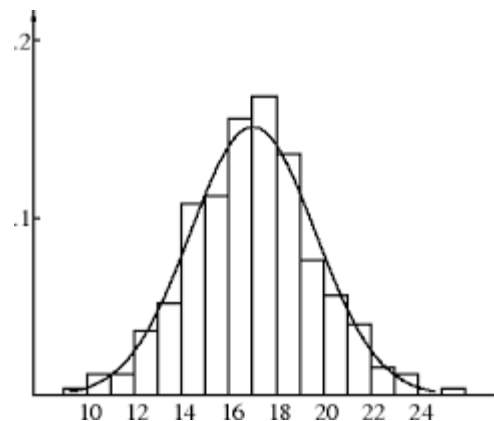
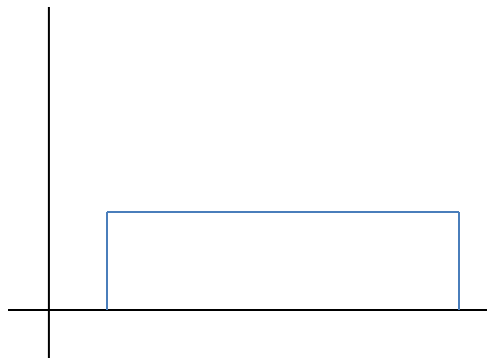
Gleichverteilung?



Normalverteilung?

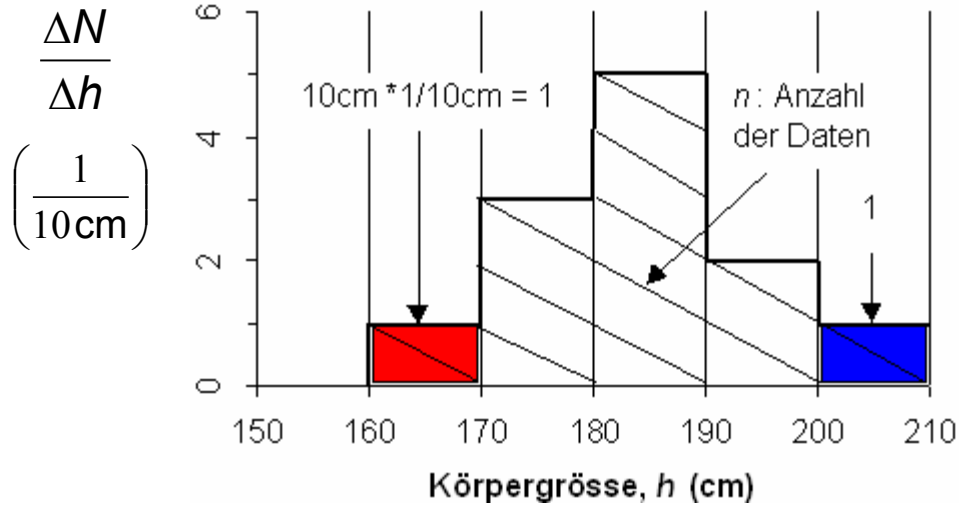


Überlagerung von zwei Normalverteilungen?



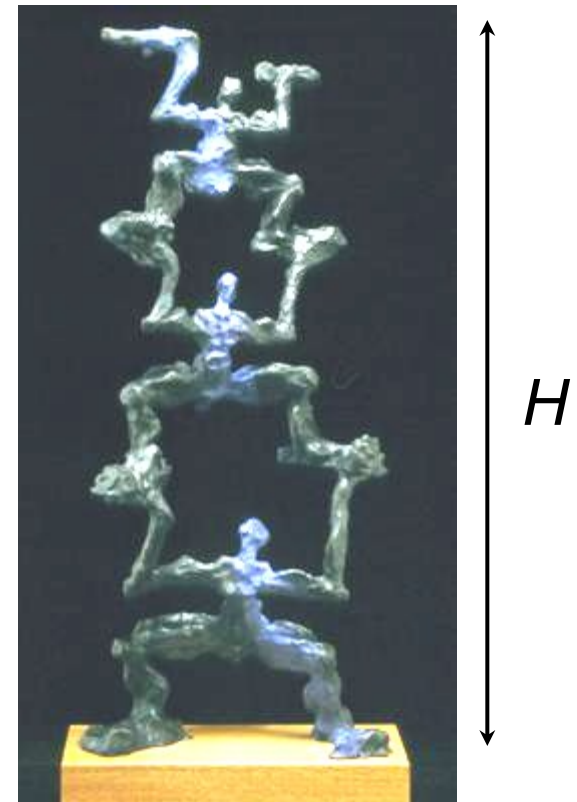
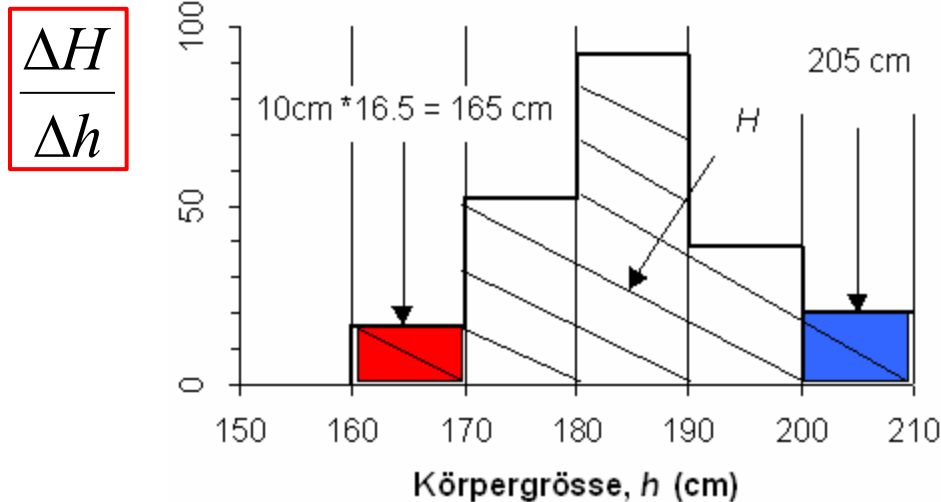
Vergleichen mit bekannten Verteilungen...

Häufigkeitsverteilung



h : Körperhöhe

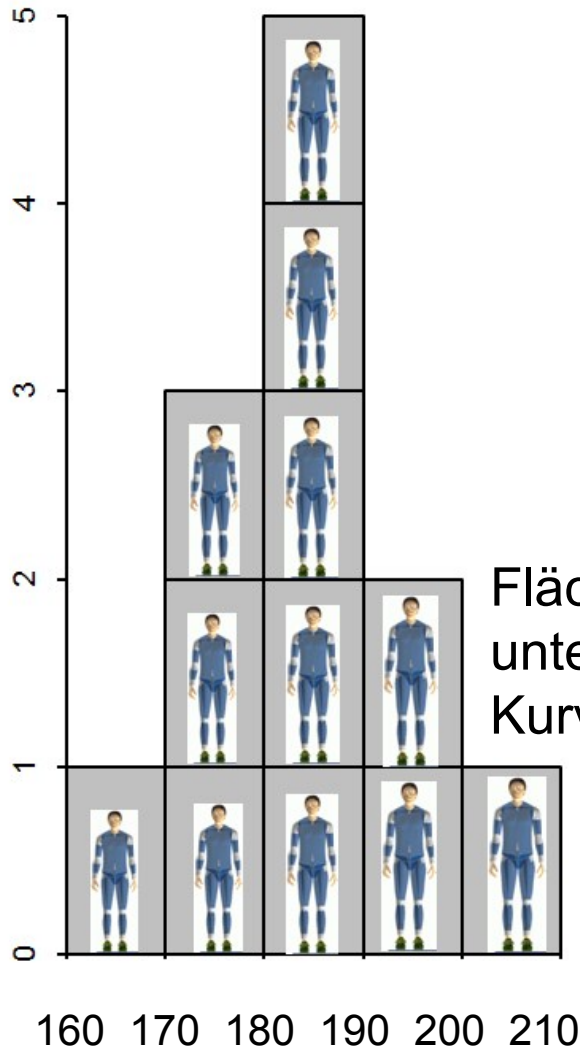
H : kollektive Höhe, Gesamthöhe



Spektrum ist eine spezielle Häufigkeitsverteilung

Häufigkeitsdichte

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



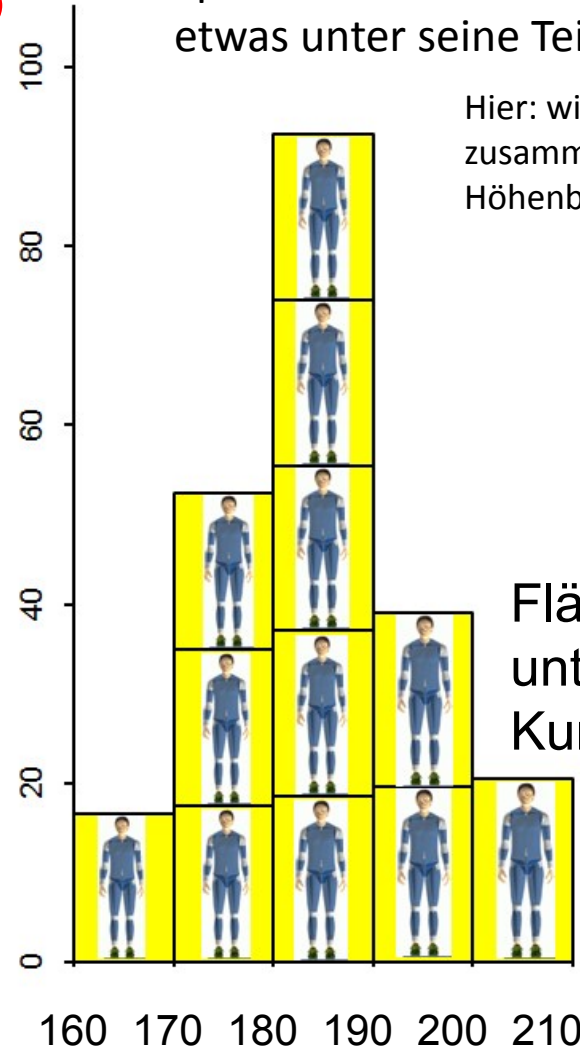
h (cm)

Spektrum

Spektrum ist die Verteilung von etwas unter seine Teile

Hier: wie viel Höhe ist zusammen in einem Höhenbereich

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

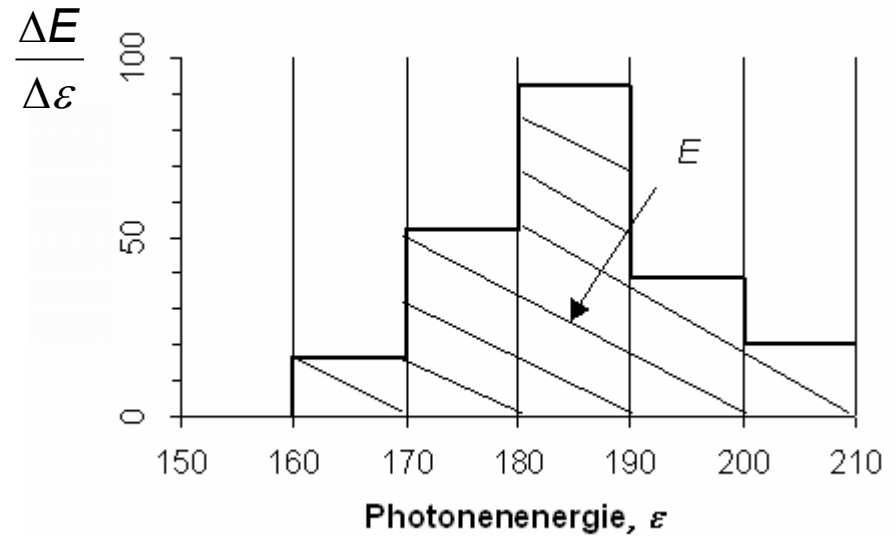


h (cm)

Beispiel aus der Physik

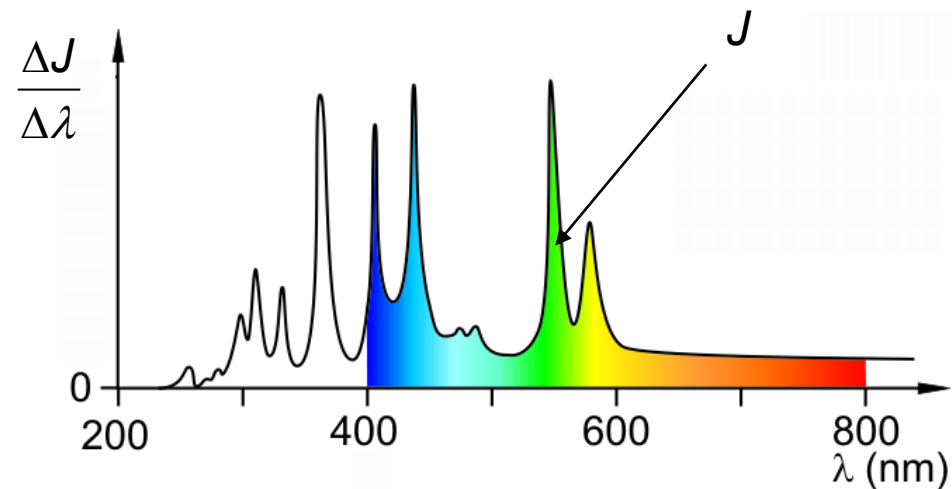
Emissionsspektrum:

wie verteilt sich die emittierte Energie über die Photonenenergien



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

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.

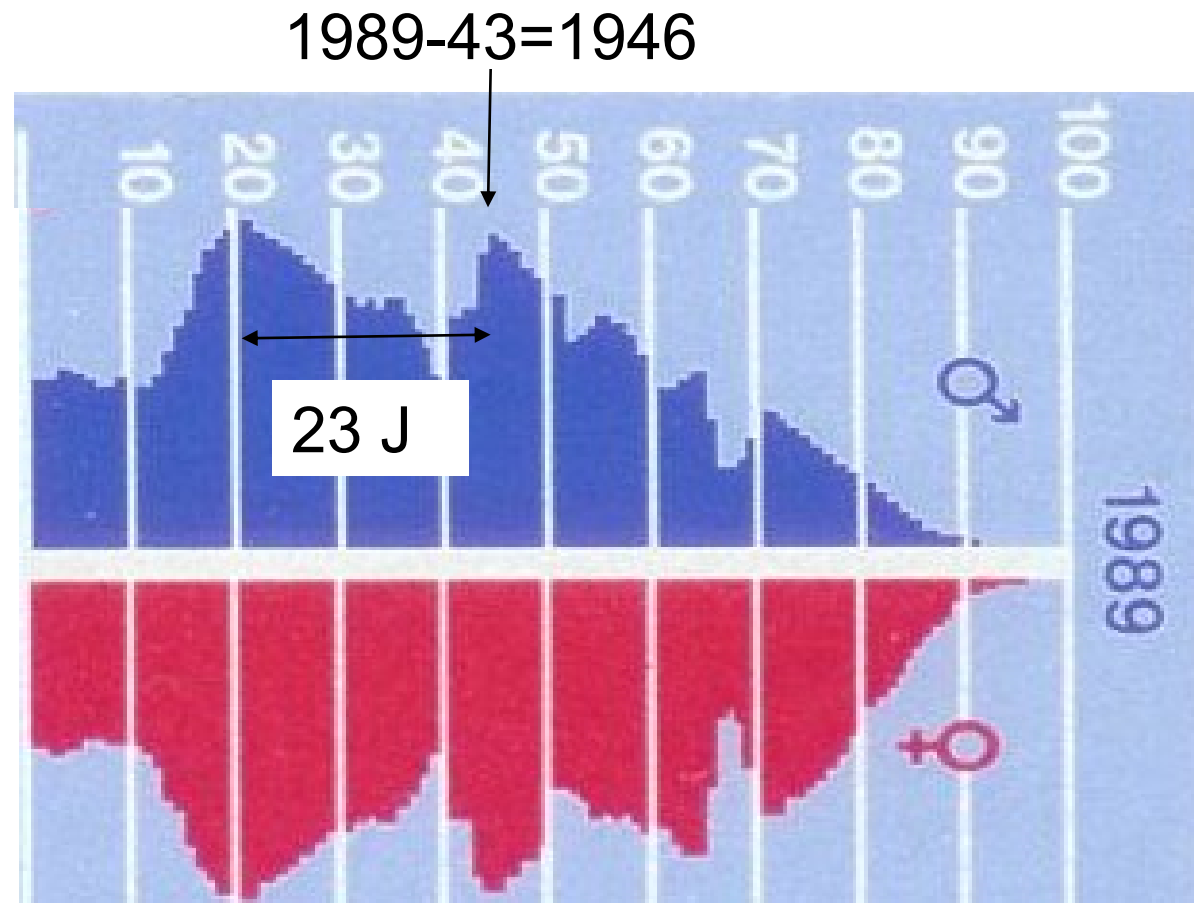
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

=Mittelwert(...)

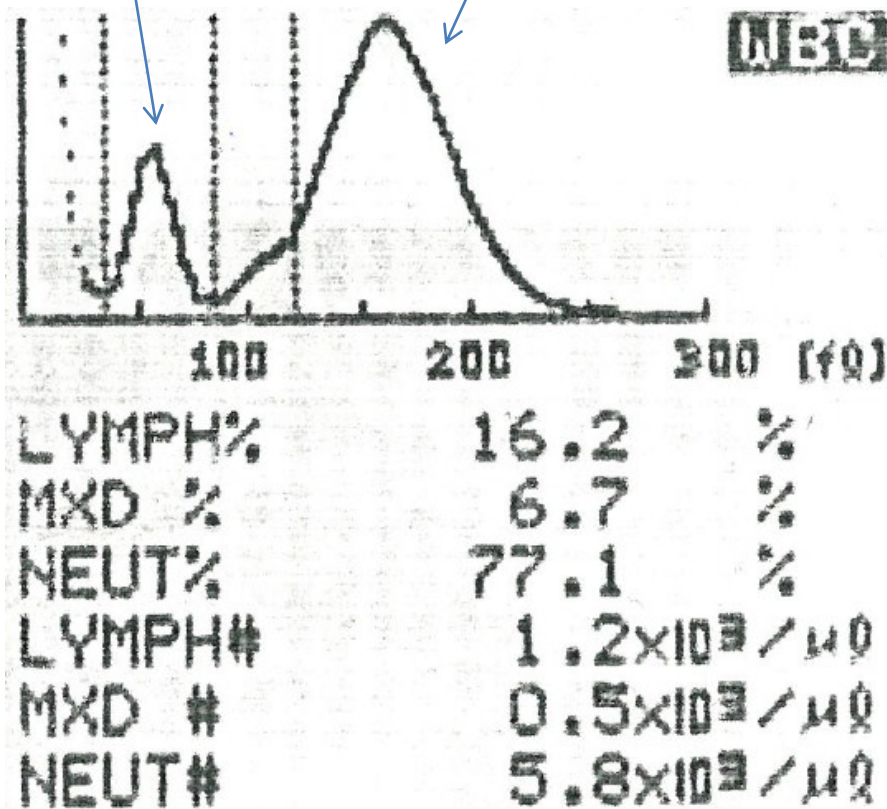
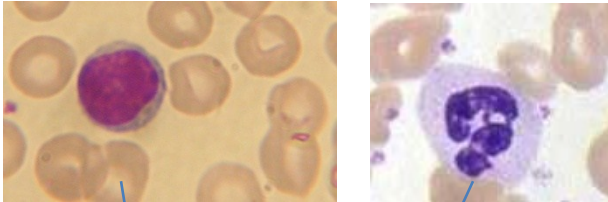
Beispiel, und modalität

Altersaufbau der deutschen Bevölkerung

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



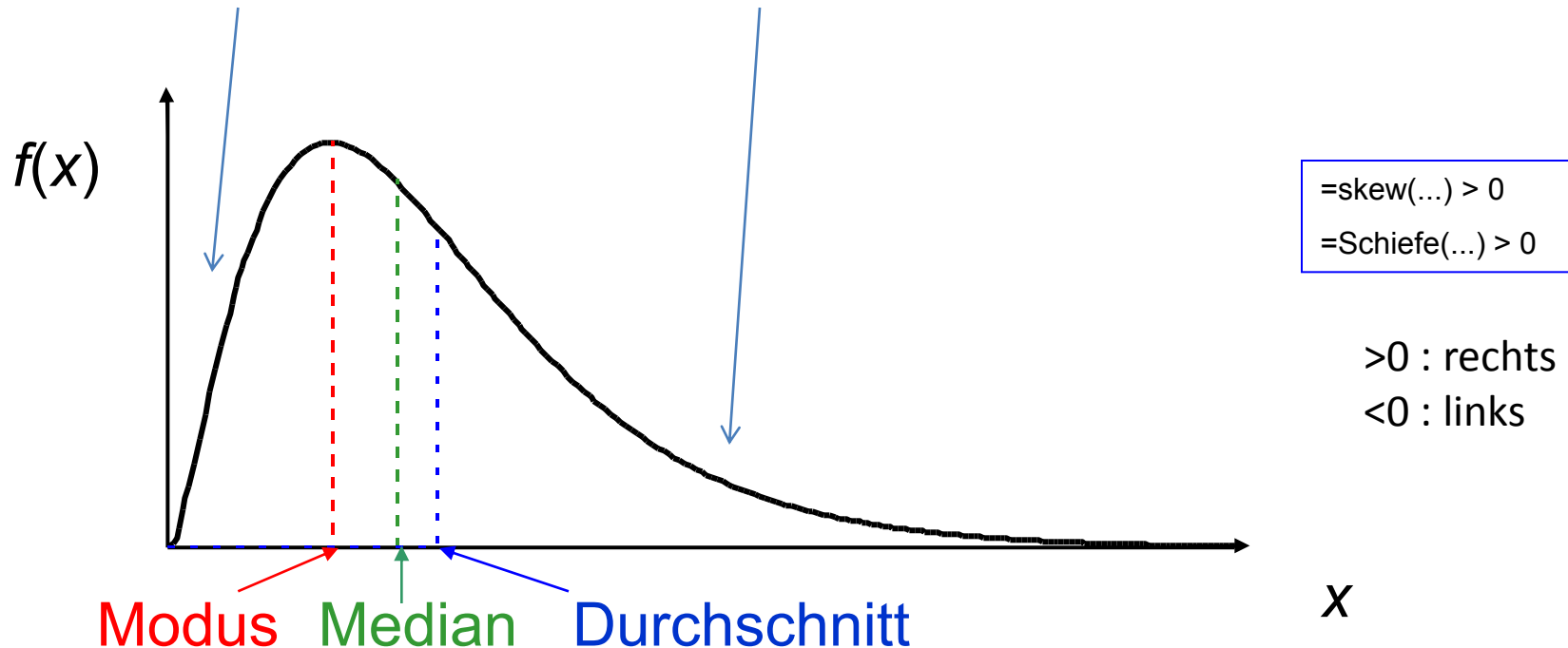
Beispiel in der Medizin



Coulter Zähler

Größenverteilung der Zellen

Linkssteile bzw. **rechtsschiefe** Verteilung

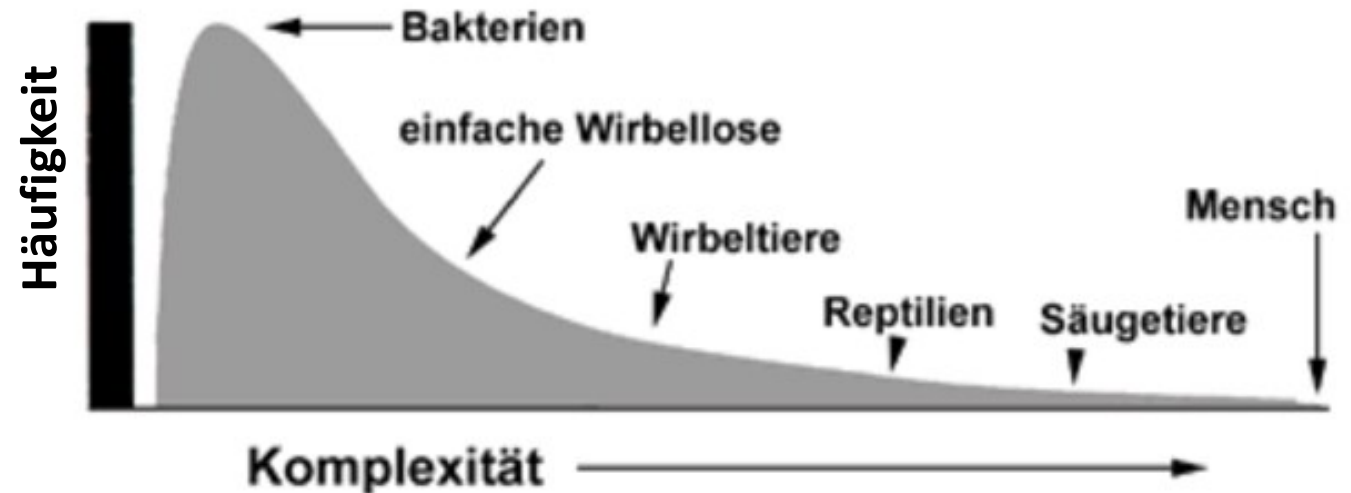
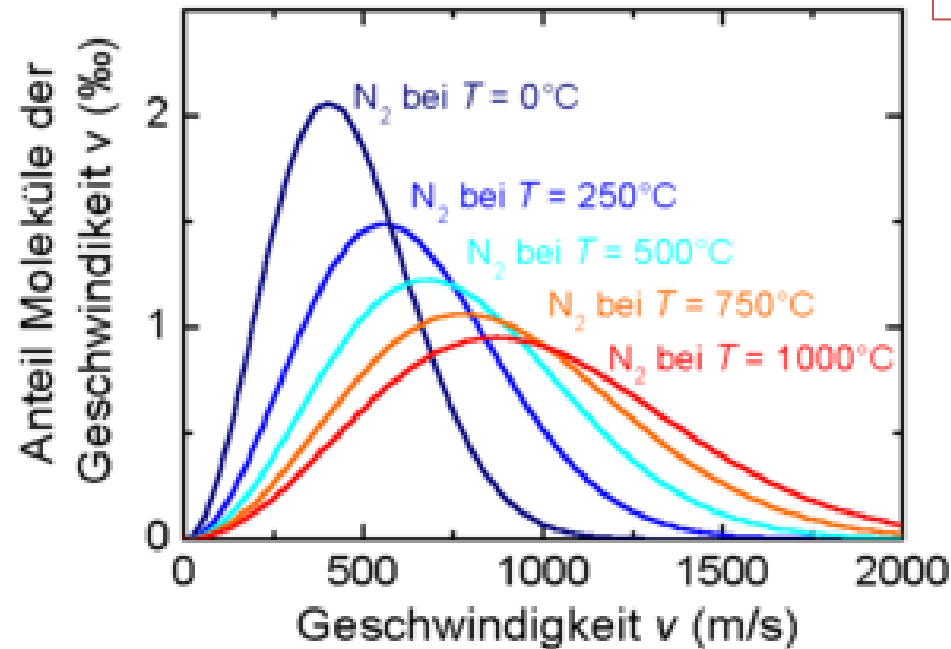


z.B. Einkommensverteilungen in einem Land:

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

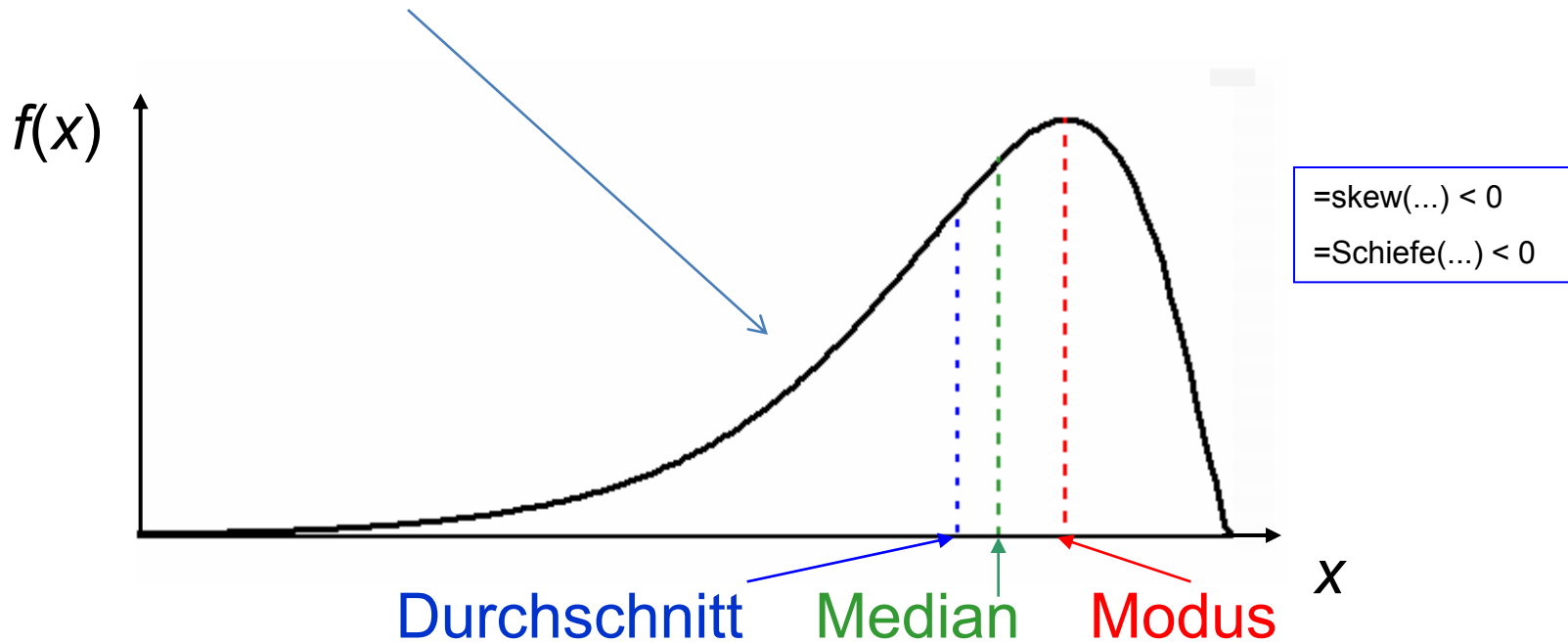
Weitere Beispiele

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

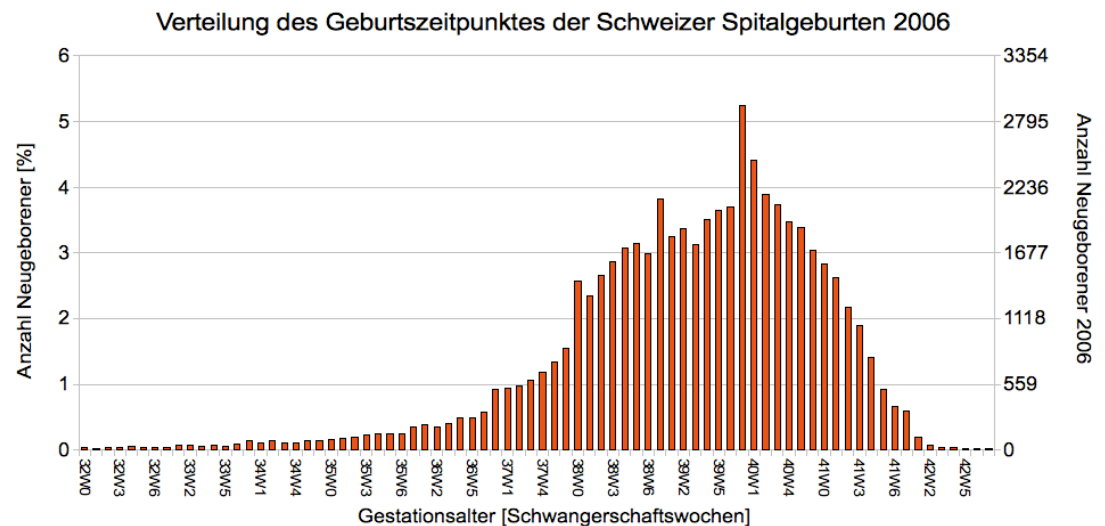


Komplexität der Tiere

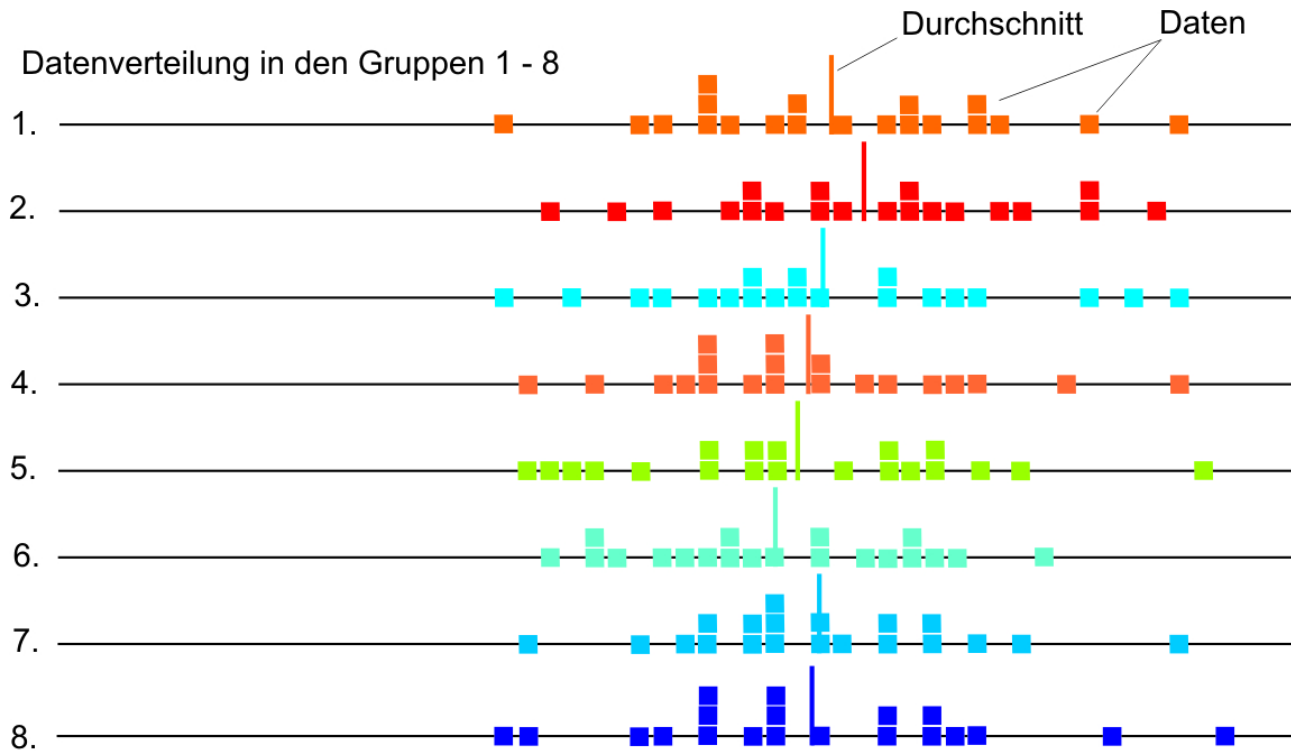
Linksschiefe bzw. rechtssteile Verteilung



z.B. Dauer einer Schwangerschaft



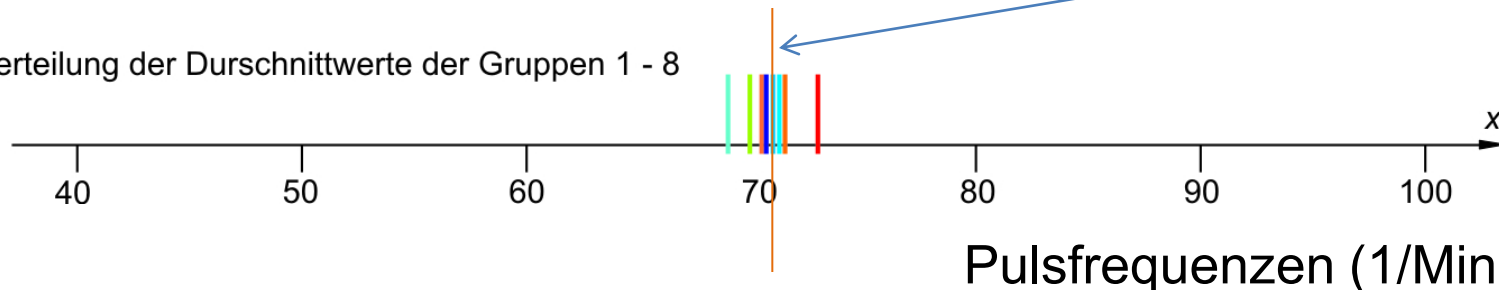
Daten und ihre Durchschnittswerte



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):

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

=Quantil(...)

Englisch: Percentile(...)

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
 α = 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.