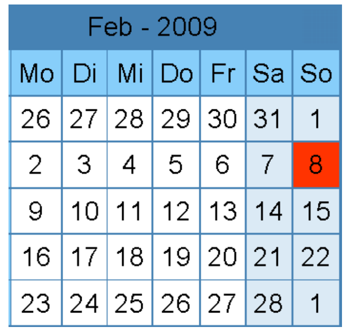
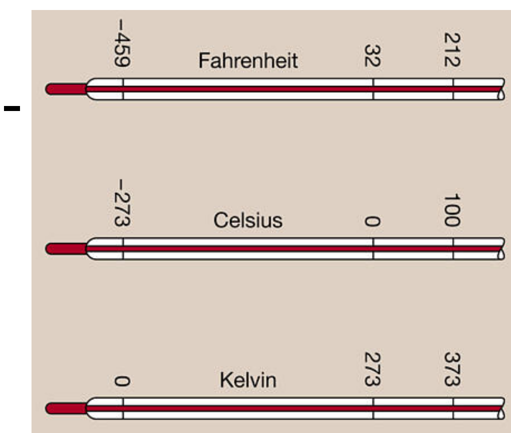
Widerholung: Klassifizierung der Merkmale

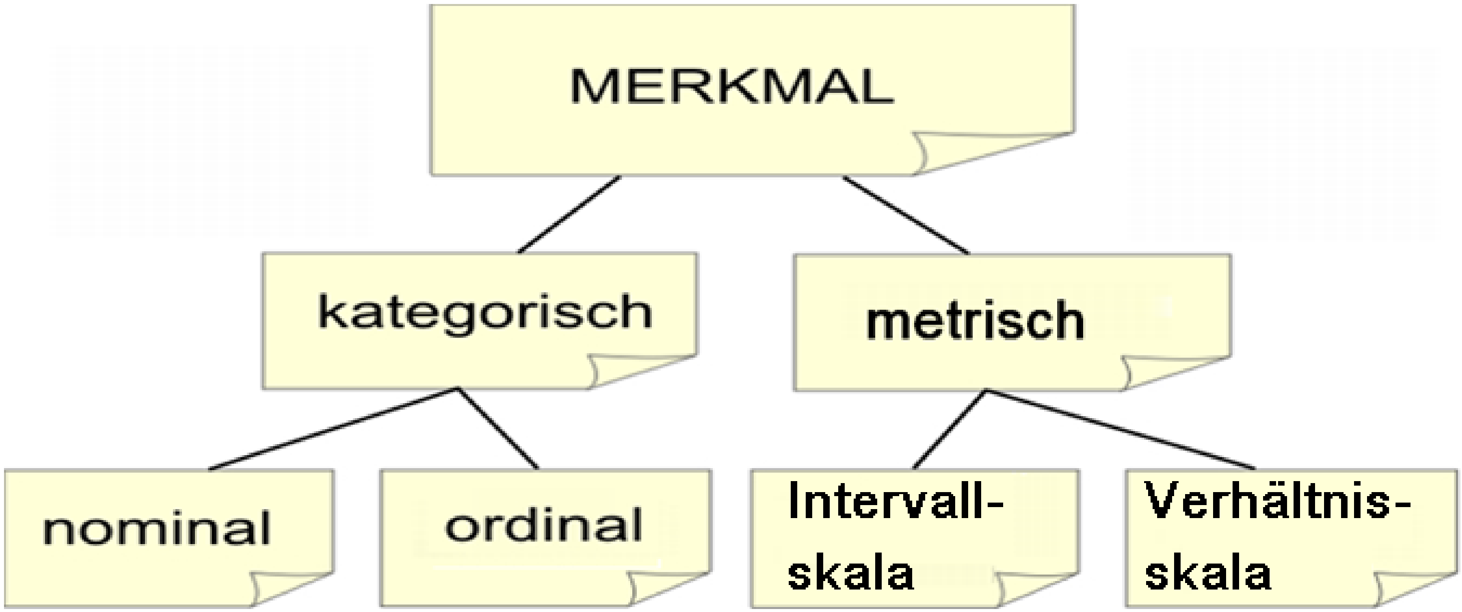
MERKMAL

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

Ab hier neu ☺

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 



$=, \neq$	$=, \neq$	$=, \neq$	$=, \neq$	**Auseinander-halten**
	$<, >$	$<, >$	$<, >$	**Anordnung**
		$+, -$	$+, -$	**Differenz**
			$*, /$	**Verhältnis**

„Enwicklungsstand" →

3

Ein Element

Grundgesamtheit (Population):

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

$N = \infty$ oder ungeheuer groβ

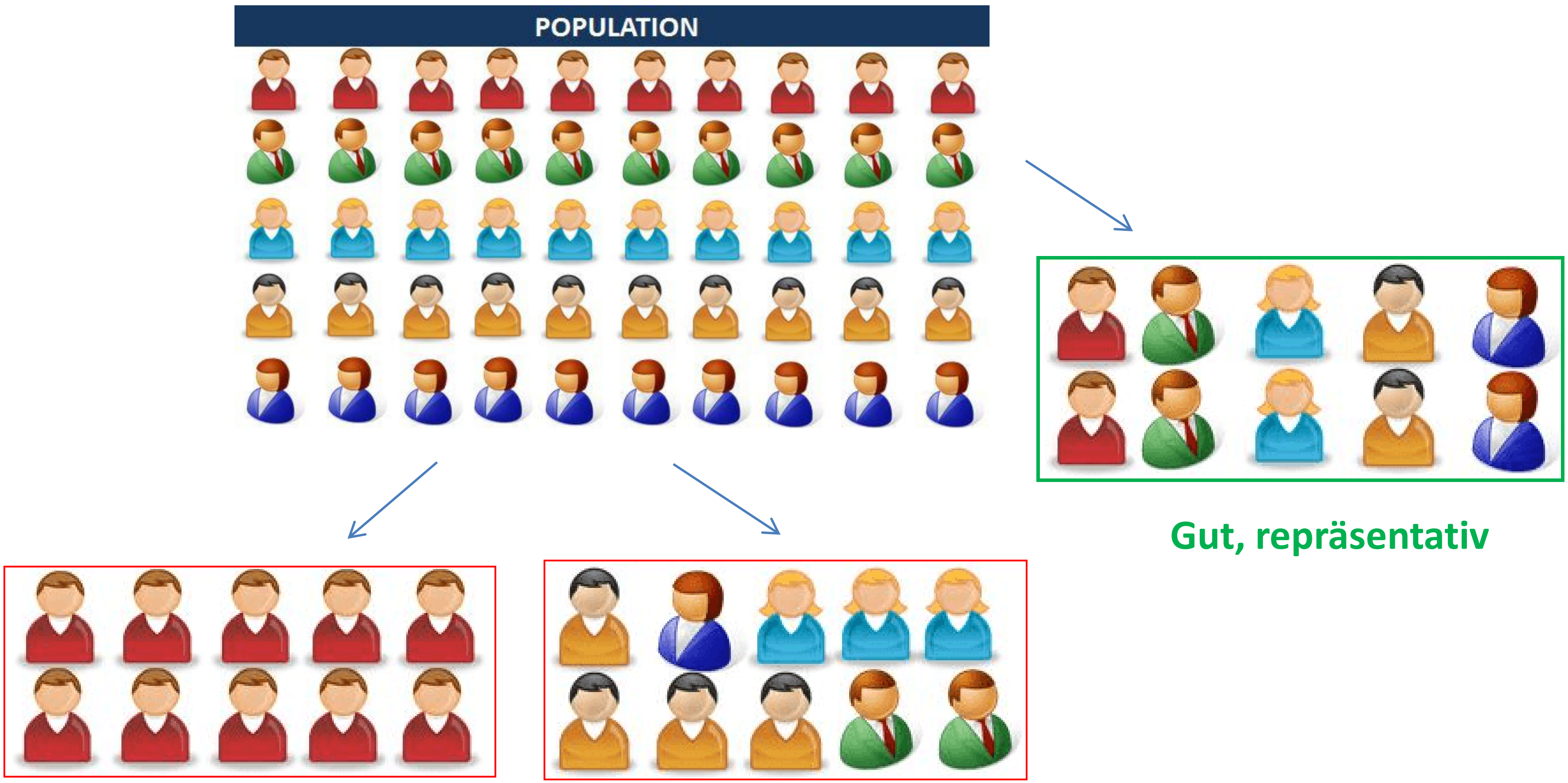
Stichprobe:

Der für die Studie ausgewählte Teil der Population.

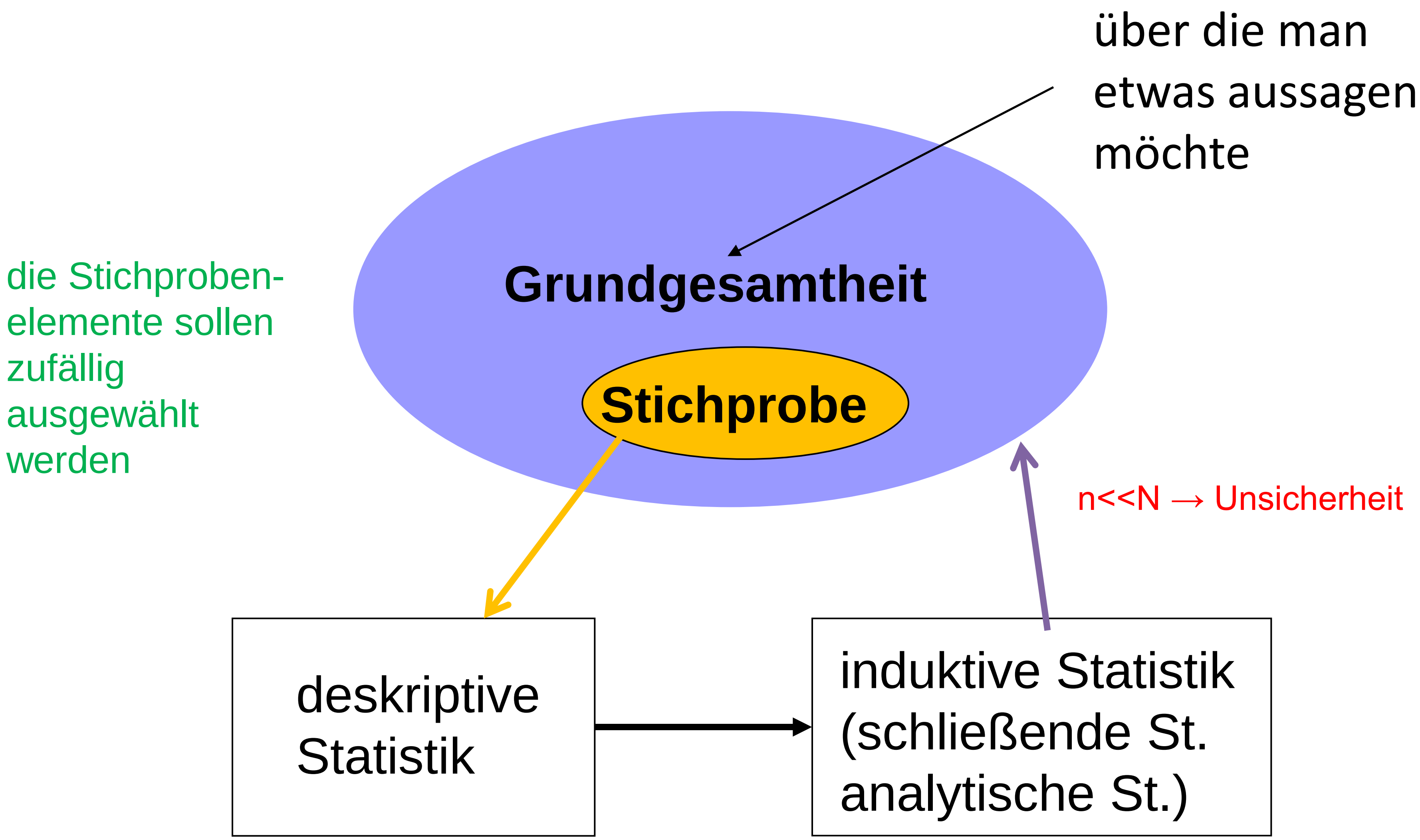
$n << N$

Umfang d. Stichprobe = Anzahl d. Daten

Wir brauchen eine **repräsentative Stichprobe**



Nicht gut!
ganz andere Zusammensetzung als die der Population


über die man
etwas aussagen
möchte
Grundgesamtheit
Stichprobe
die Stichproben-
elemente sollen
zufällig
ausgewählt
werden
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

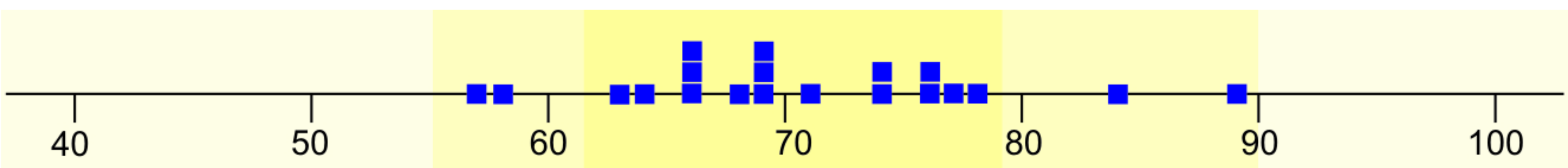
66	56	89	63	66	69	71	68	58	69
78	66	64	84	74	76	69	77	74	76

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.

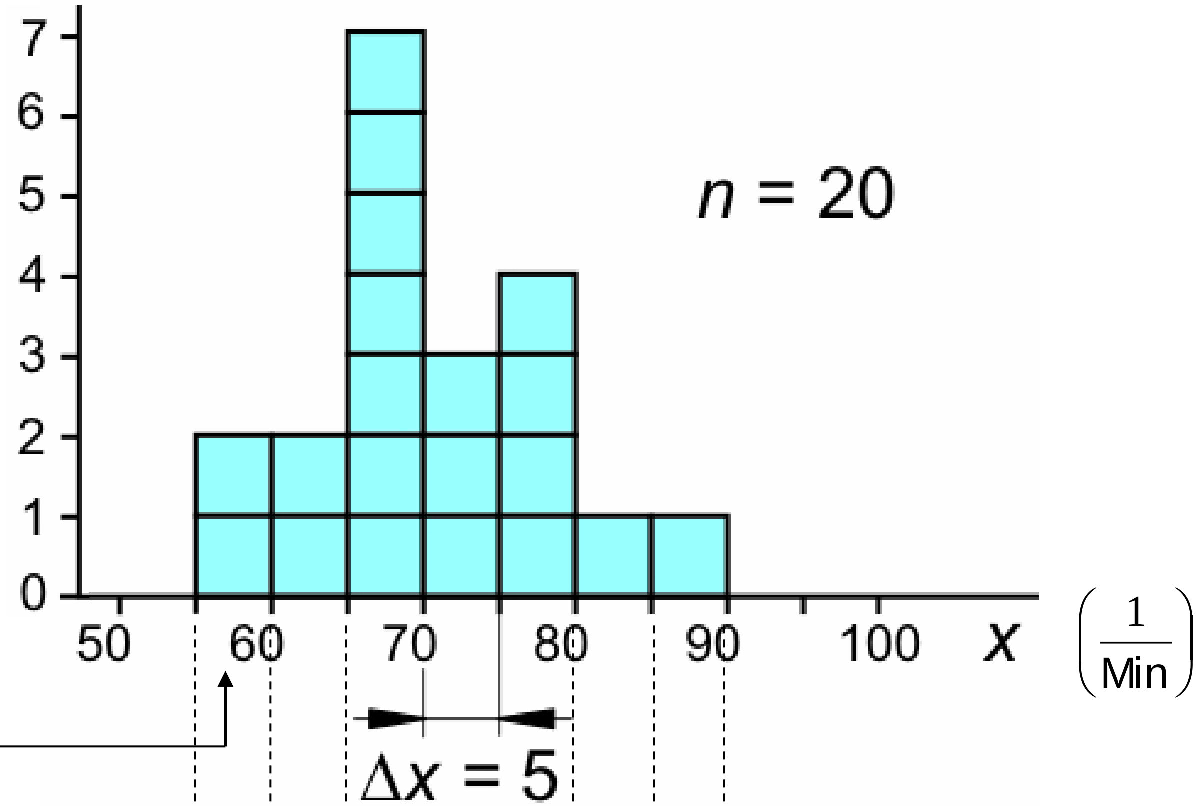
Pr.Buch Tabelle 5

Häufigkeitsdichte

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

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

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



Die Fläche unter der Treppenfunktion zwischen 55 und 60: $5\frac{1}{\text{Min}} \cdot 2\frac{\text{Min}}{5} = 2$

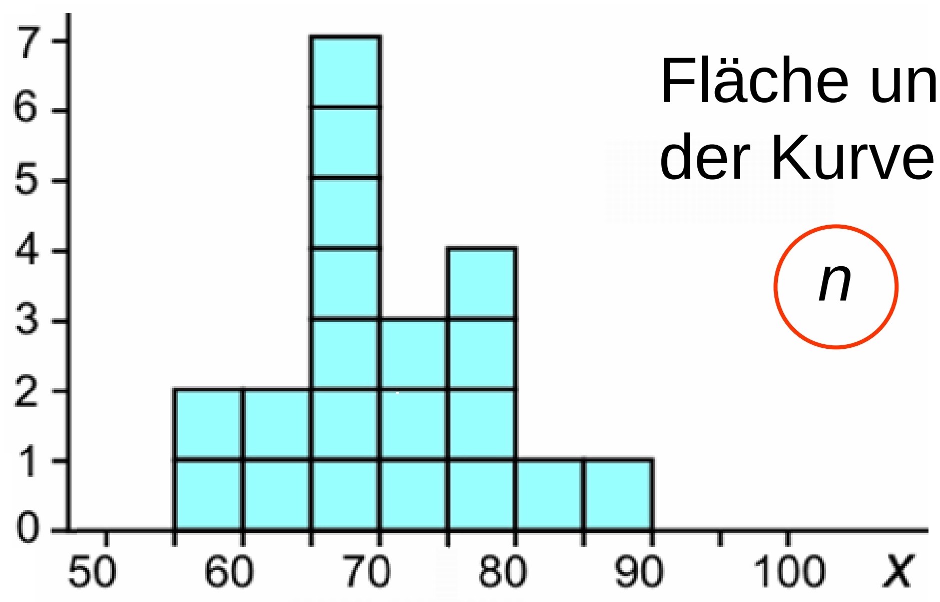
KLASSENGRENZEN	HÄUFIGKEIT
$55 \leq x_i < 60$	2
$60 \leq x_i < 65$	2
$65 \leq x_i < 70$	7
$70 \leq x_i < 75$	3
$75 \leq x_i < 80$	4
$80 \leq x_i < 85$	1
$85 \leq x_i < 90$	1
insgesamt:	$n = 20$

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



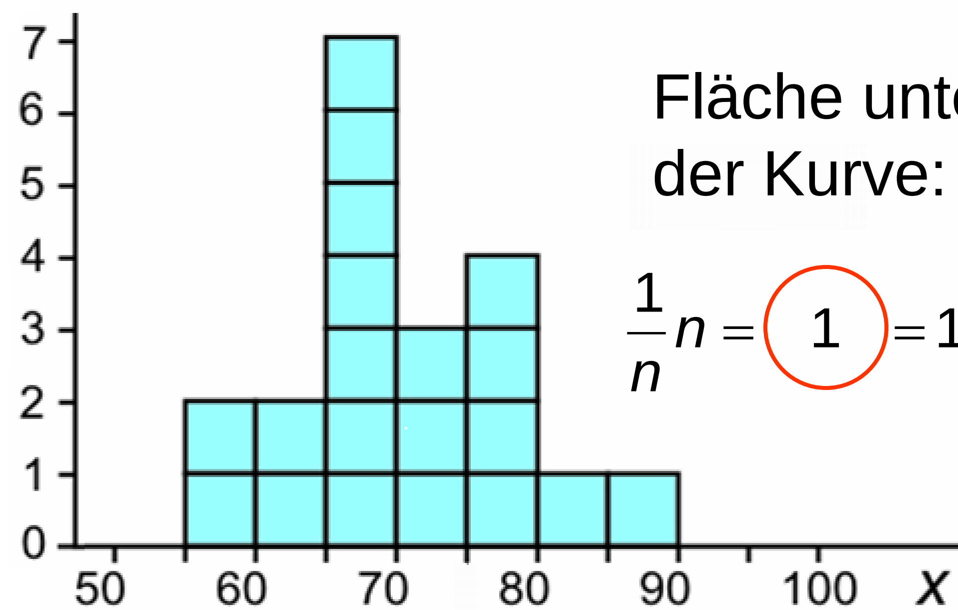
Fläche unter der Kurve:

n

Häufigkeitsdichte-verteilung

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

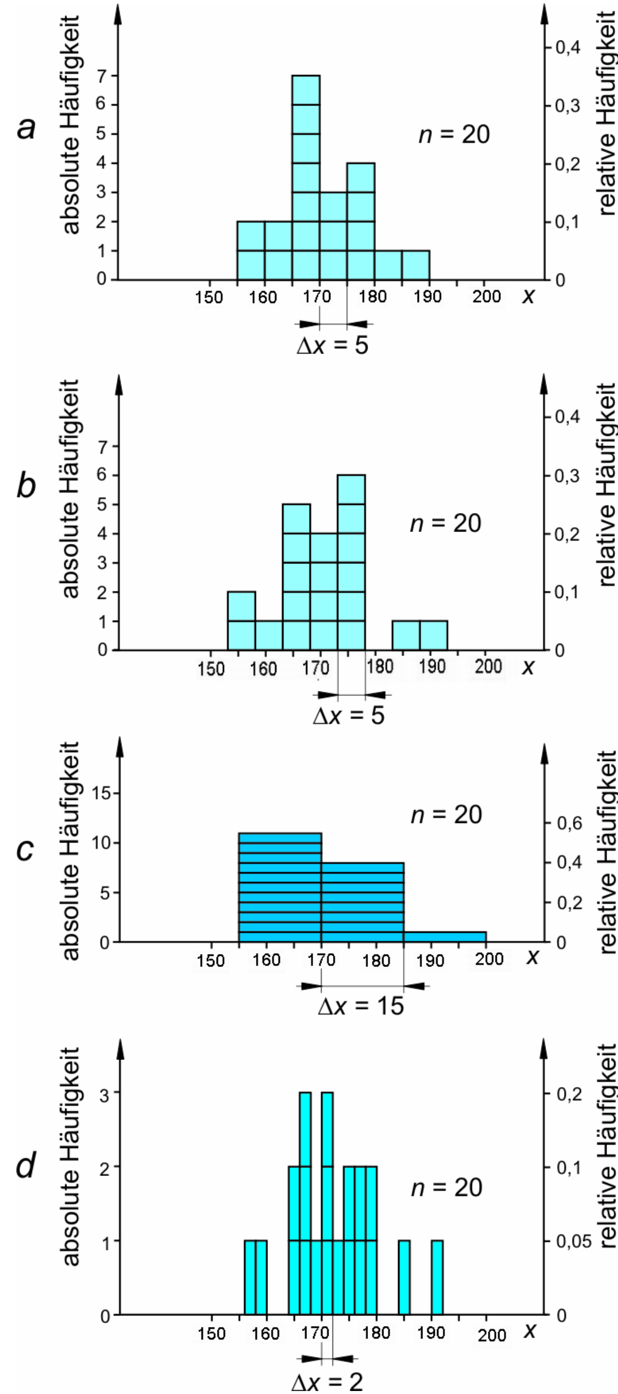
relative



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

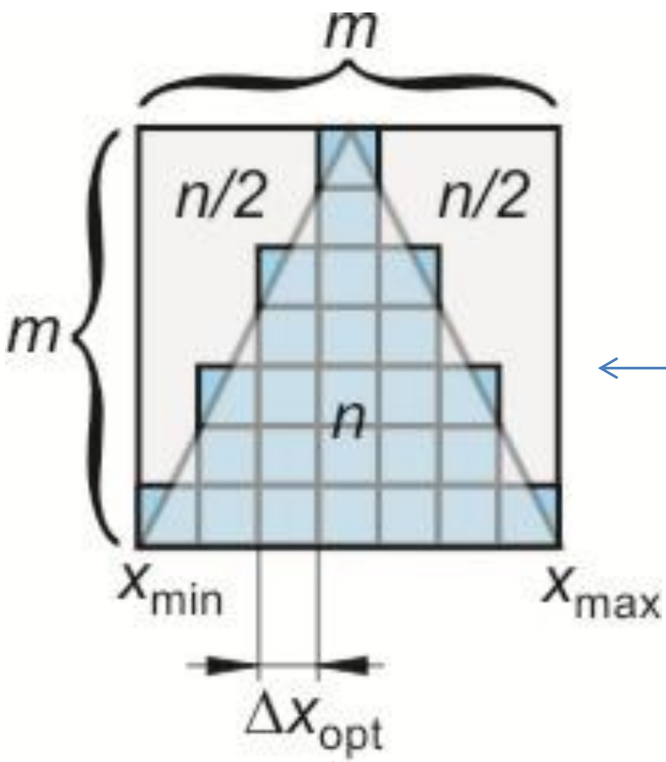


Selbe Grundgesamtheit, 2 Stchproben

Zu groβe 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*:

$$m^2 = 2n$$

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

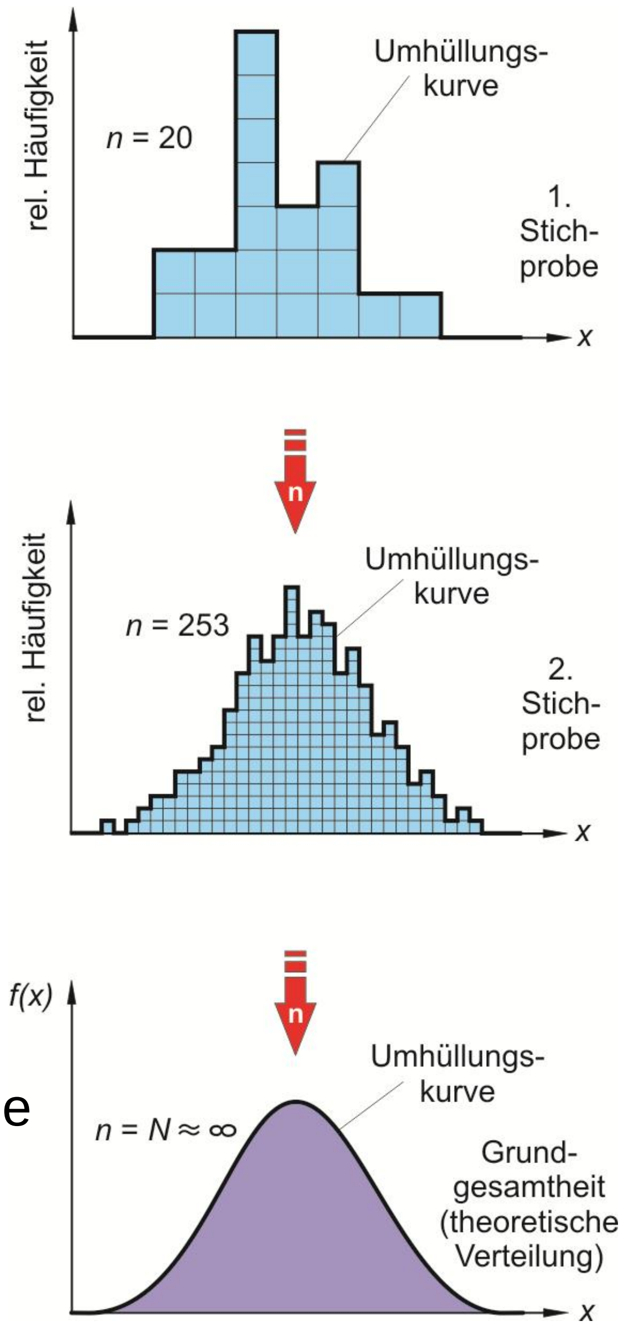
optimale Klassenbreite Δx:

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

empirische
Funktion

empirische
Funktion

theoretische
Funktion



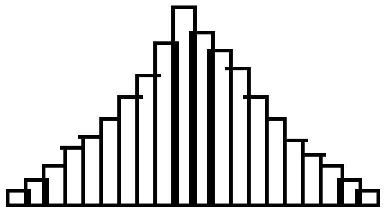
Pr.Buch Abb. 6

n vergrößert sich,
die Klassenbreite Δx kann
verkleinert werden

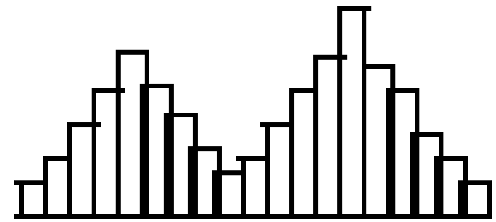
Bei großen Stichproben ergibt die empirische Verteilungsfunktion **eine sehr gute Näherung** der theoretischen Verteilungsfunktion. (Die Stichprobe ist „fast gleich" der Grundgesamtheit.)

Analyse von Häufigkeitsverteilungen

homogene symmetrische Stichprobe:

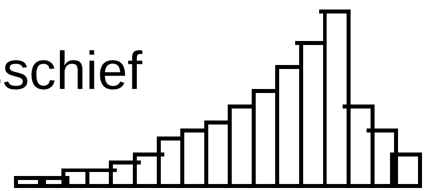


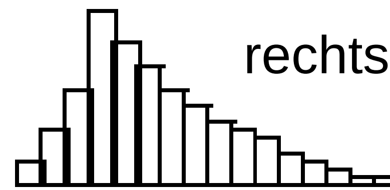
heterogene Stichprobe:



homogene nichtsymmetrische Stichproben:

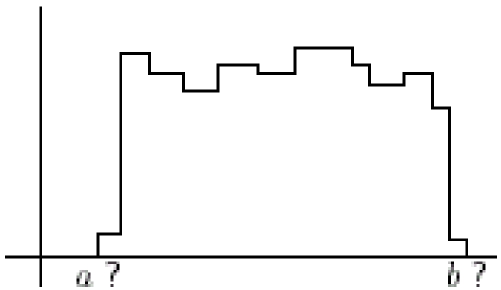
linksschief



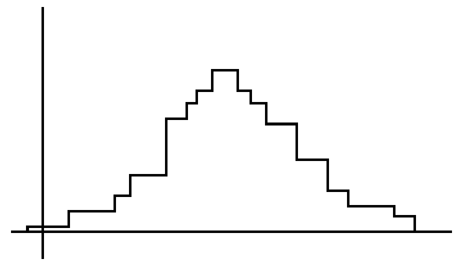


rechtsschief

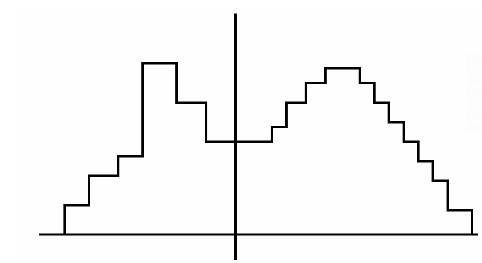
Vermutungen macht man auch:



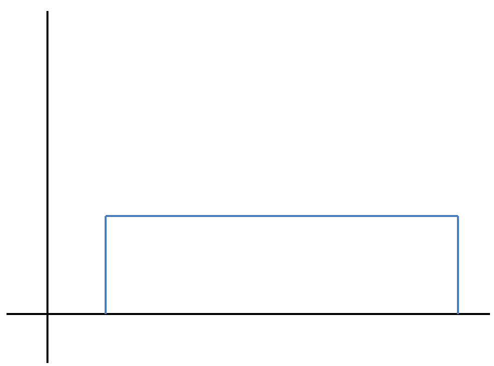
Gleichverteilung?

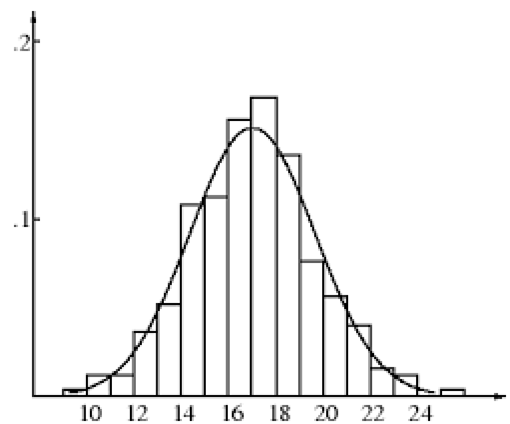


Normalverteilung?



Überlagerung von zwei Normalverteilungen?





Vergleichen mit bekannten Verteilungen...