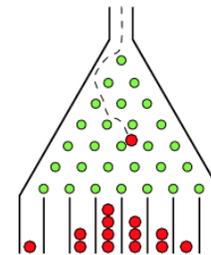




## Deskriptive Statistik 1

KAD 2012.09.13



Die Statistik beschäftigt sich mit **Massenerscheinungen**, bei denen die dahinterstehenden Einzelereignisse meist zufällig sind.

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.

2

## 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 Anwendungsgebiet der Statistik bestand in der **Staatsbeschreibung** (Völkzählung)  
*Status = Zustand*



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



4

## Was messen Physiker, Arzt und Medizinstudent?

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

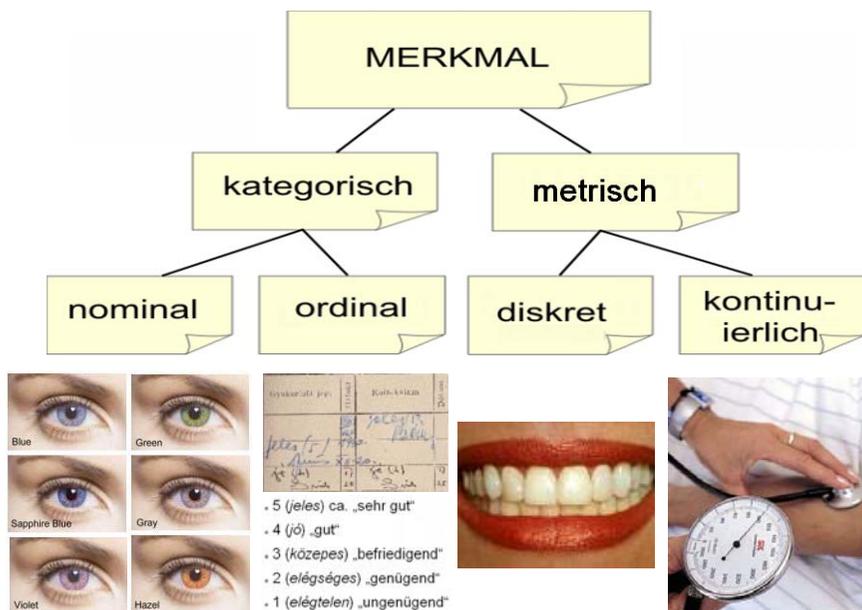
5

Pr.Buch 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. BURGENDORFERI-AK (EIA) IGM		positiv	positiv	positiv	positiv			
B. BURGENDORFERI-AK IGG (EIA)		negativ	negativ	negativ	negativ	5	10	
Ery.-Vert.-Breite	%		11,6			11,6	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	16,0	
Leukozyten	/ul		7			6,5	10,0	
MCH	pg		32,1			32,1	34,0	
MCHC	g/dl		34,0			34,0	37,0	
MCV	ucm		94,4			94,4	99,0	
P 18 (p18-Protein)		negativ	negativ	negativ	negativ			

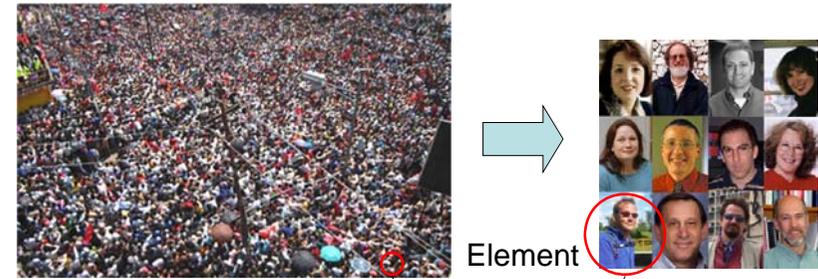
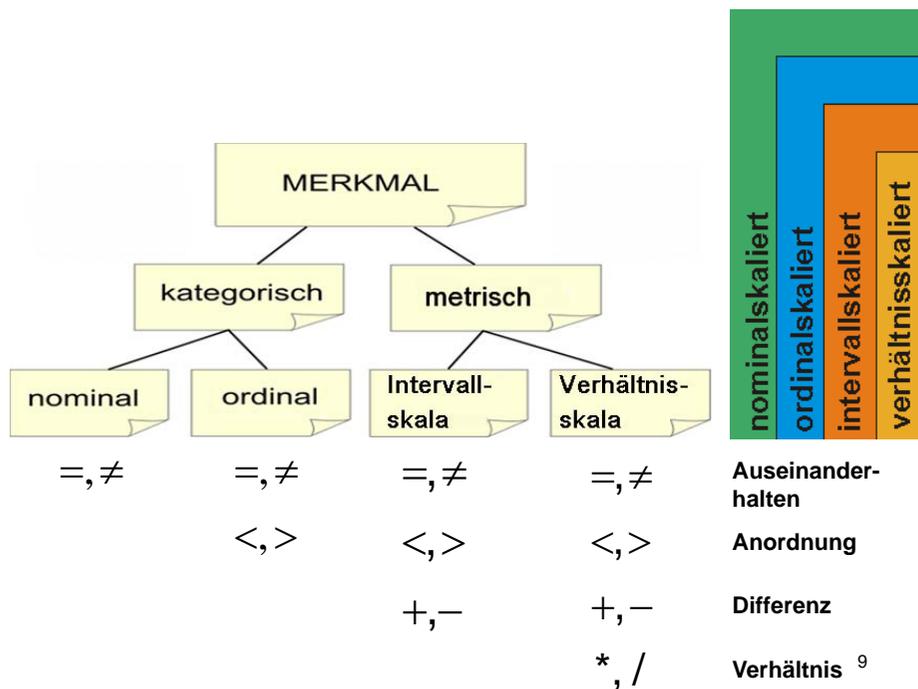
## Klassifizierung der Merkmale



## Skalentypen der numerischen Merkmale

	diskret	kontinuierlich
<b>Intervallskala</b> definierte Differenz, „kein“ 0 Punkt	Tage in einem Kalender 	Temperatur in °C 
<b>Verhältnisskala</b> definiertes Verhältnis, 0 Punkt	Anzahl der Zähne 	Temperatur in K 

8



**Grundgesamtheit (Population):**

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

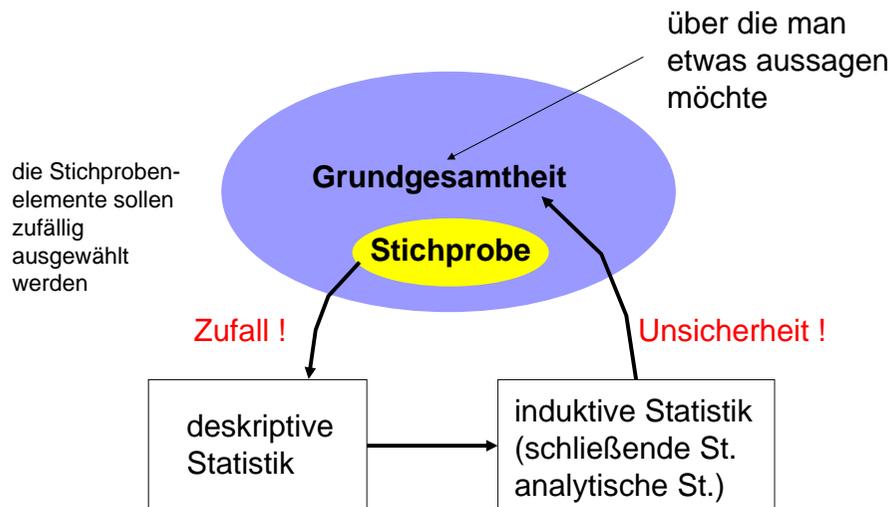
$N = \text{„unendlich“}$

**Stichprobe:**

Der für die Studie ausgewählte Teil der Population.

$n = \text{endlich}$

$N \gg n$  (Umfang)



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**

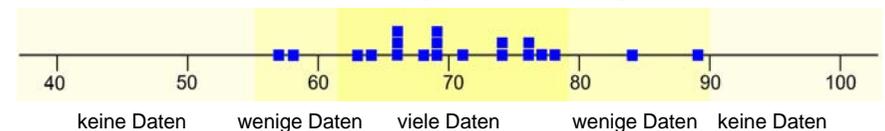
Daten der Stichprobe liegen in Form einer Urliste vor:

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!



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!

KLASSENRENZEN	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
<b>insgesamt:</b>	$n = 20$

in Excel:

=frequency(...)  
=Häufigkeit(...)

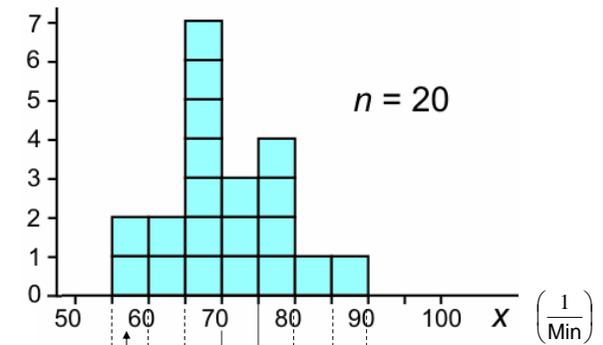
Die Grenzwerte und die Breiten der Klassen sind willkürlich. Stellen wir diese Treppenfunktion dar!

13

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

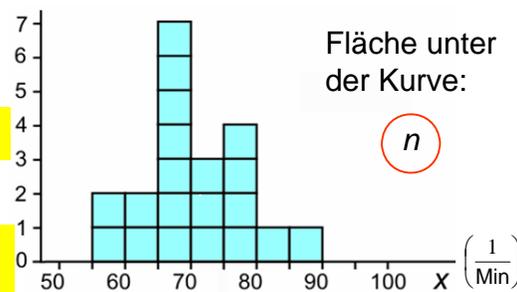
KLASSENRENZEN	HÄUFIGKEIT
$55 \leq x_i < 60$	2
$60 \leq x_i < 65$	2
$65 \leq x_i < 70$	7
$70 \leq x_i < 75$	3
$75 \leq x_i < 80$	4
$80 \leq x_i < 85$	1
$85 \leq x_i < 90$	1
<b>insgesamt:</b>	$n = 20$

14

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

**absolute**

**Häufigkeitsdichte-  
verteilung**

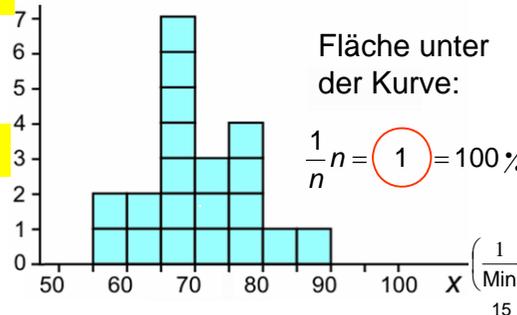


Fläche unter der Kurve:

$n$

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

**relative**

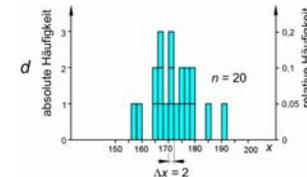
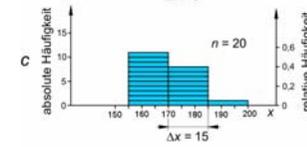
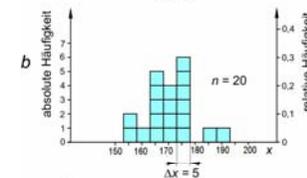
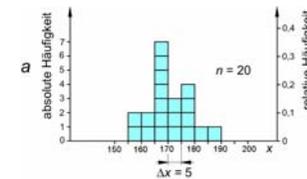


Fläche unter der Kurve:

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

15

**absolute  
Häufigkeits-  
dichte  
(Histogramm)**



**relative  
Häufigkeits-  
dichte  
(Histogramm)**



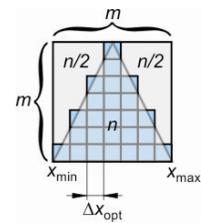
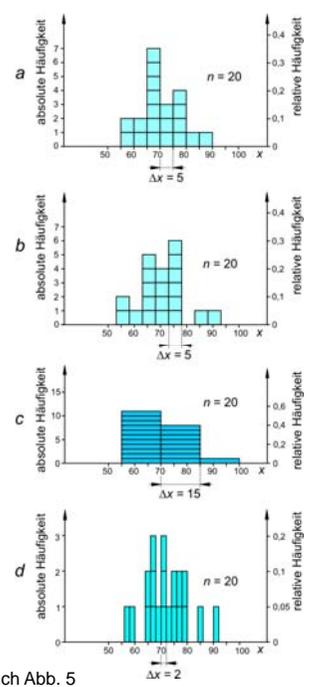
„Jedes Rechteck entspricht einem Messwert.“



$n$  vergrößert sich, die Klassenbreite  $\Delta x$  kann verkleinert werden

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

## Bestimmung der optimalen Klasseneinteilung



optimale Klassenanzahl  $m$ :

$$m^2 = 2n$$

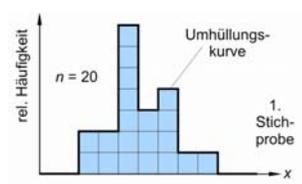
$$m = \sqrt{2n} \quad m = \sqrt{40} = 6.3$$

optimale Klassenbreite  $\Delta x$ :

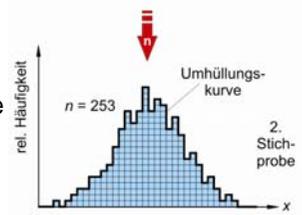
$$\Delta x = \frac{x_{\max} - x_{\min}}{m} \quad \Delta x = \frac{89 - 56}{6.3} = 5.2$$

Pr.Buch Abb. 5

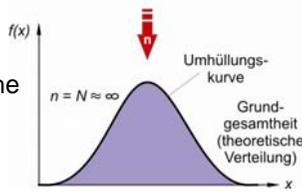
empirische Funktion



empirische Funktion



theoretische Funktion



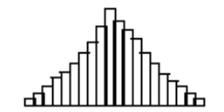
Pr.Buch Abb. 6

## Beispiel: Biophysik Praktikum, Mikroskop

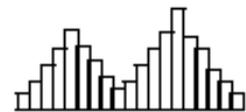
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	calibr.				total		N1 2009.09.23	Mori	Pia+	Isab	Jan	Stef	Fre	Sim	Meli	Jan	Nata	
2	um/o.u.	2.5			um	o.u.	unit	o.u.										
3	confidence int.				22	8.8	average	10	8.4	8.3	8.6	8.6	9.1	8.4	9.1	8.8	8.4	
4	95% data, um		95% data, o.u.		3.68	1.47	stdev	1.5	1.4	1.3	1.1	1.9	1	1.2	1.3	1.6	1.3	
5	14.6	29.4	5.9	11.7	0.18	0.07	s.e.m	0.2	0.2	0.2	0.2	0.3	0.1	0.2	0.2	0.2	0.2	
6	95% avg, um		95% avg, o.u.		7.5	3	min	6	6	4	6	3	7	6	6.5	5	5	
7	21.6	22.4	8.7	8.9	37.5	15	max	15	11	10	12	15	11	11	12	12	11	
8								1	2	3	4	5	6	7	8	9	10	
9								1	10	10	9	9	10	10	8	11	8	10
10								2	9.5	9	7	8	9	9	9	11	8	
11								3	10	10	7	10	10	9	10	8	9	9
12								4	9	8	8	9	10	8	9	9	8	10
13								5	10	7	10	8	8	9	7	9.5	10	8
14								6	10	9	6	8	11	9	8	11	9	8
15								7	10	9	7	8	8	9	9	9.5	10	7
16								8	11	8	8	10	11	11	10	10	12	8
17								9	11	8	9	9	10	8	8	9	9	9
18								10	6	6	8	8	8	9	9	10	10	9
19								11	10	9	9	7	10	10	8	11	10	10
20								12	11	11	9	8	8	8	11	10	8	6
21								13	11	7	7	8	10	9	8	11	11	9
22								14	9	9	9	9	9	8	11	7	9	8
23								15	11	9	6	9	11	9	8	11	7	8
24								16	12	8	4	10	15	10	10	9	10	5
25								17	10	10	7	9.5	8	9	9	10	10	10
26								18	8	7	10	9.5	9	7	8	10	9	9
27	bin				fran	fral	fr	10	11	0	8	0	8	7	0	0	0	10

## Analyse von Häufigkeitsverteilungen

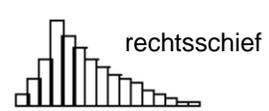
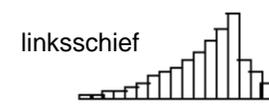
homogene symmetrische Stichprobe:



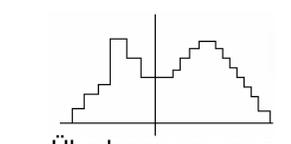
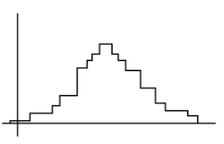
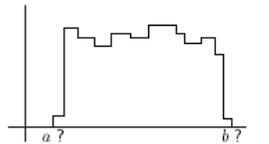
heterogene Stichprobe:



homogene nichtsymmetrische Stichproben:



Vermutung:



Gleichverteilung?

Normalverteilung?

Überlagerung von zwei Normalverteilungen?