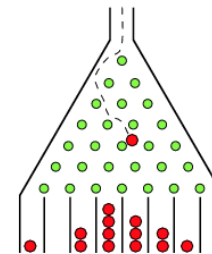




Deskriptive Statistik 1

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



das erste Anwendungsgebiet 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



4

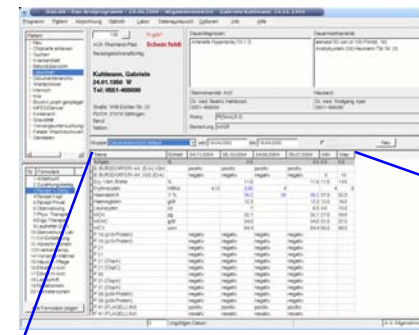
Was messen Physiker, Arzt und Medizinstudent?

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

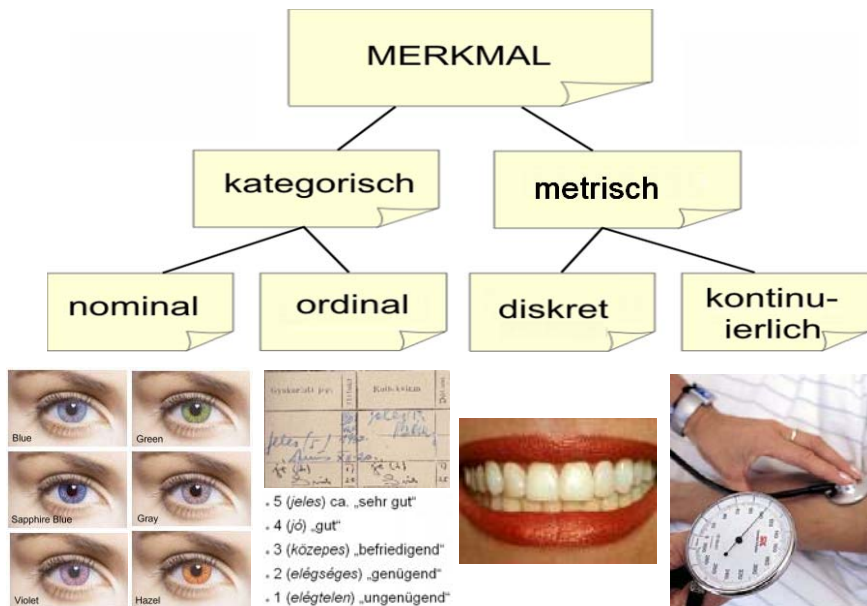
Pr.Buch Tabelle 3

Labormessergebnisse



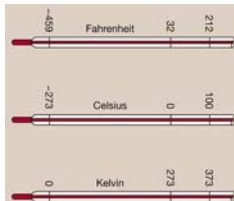


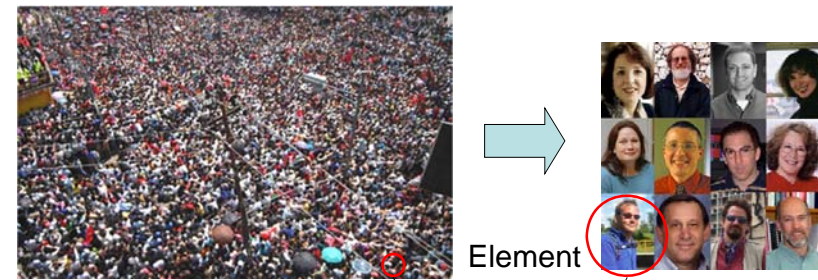
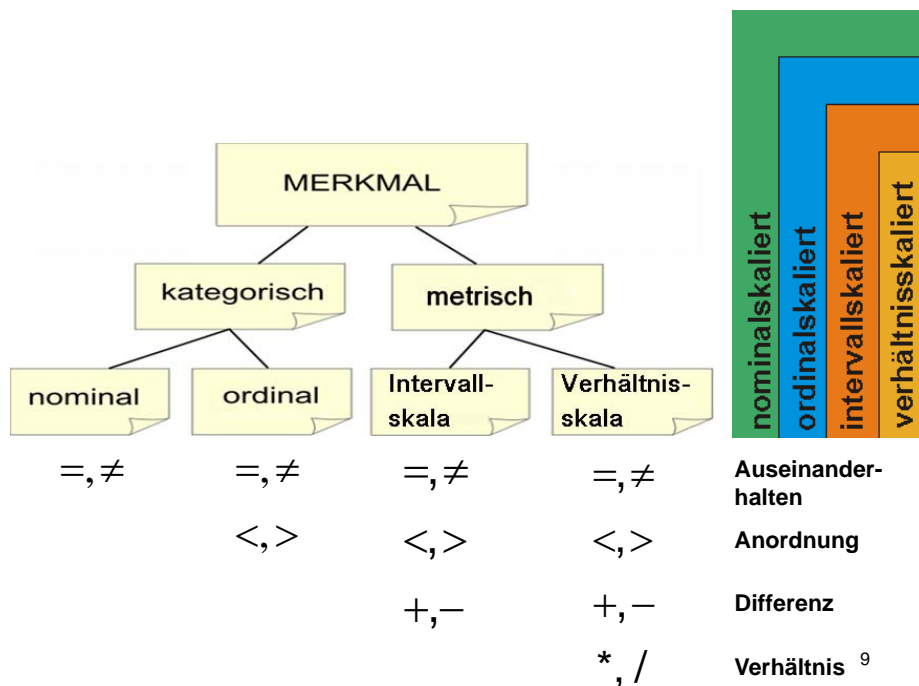
Name	Einheit	04.11.2004	05.10.2004	04.08.2004	05.07.2004	Min	Max
%Hypo	%		0.5		0.5	0.0	5.0
B. BURGENDORFERI-AK (EIA) IGM		positiv	positiv	positiv	positiv		
B. BURGENDORFERI-AK IGG (EIA)		negativ	negativ	negativ	negativ	5	10
Ery.-Vert.-Breite	%		11.6		11.6	11.5	14.5
Erythrozyten	Milliul	4,12	3,95	4		4	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		

Klassifizierung der Merkmale



Skalentypen der numerischen Merkmale

	diskret	kontinuierlich																																																	
Intervall-skala definierte Differenz, „kein“ 0 Punkt	Tage in einem Kalender <table border="1"><thead><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></thead><tbody><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></tbody></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	Tempe- ratur in °C 
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																																													
Verhältnis-skala definiertes Verhältnis, 0 Punkt	Anzahl der Zähne 	Tempe- ratur in K 																																																	



Grundgesamtheit (Population):

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

$N = \text{„unendlich“}$

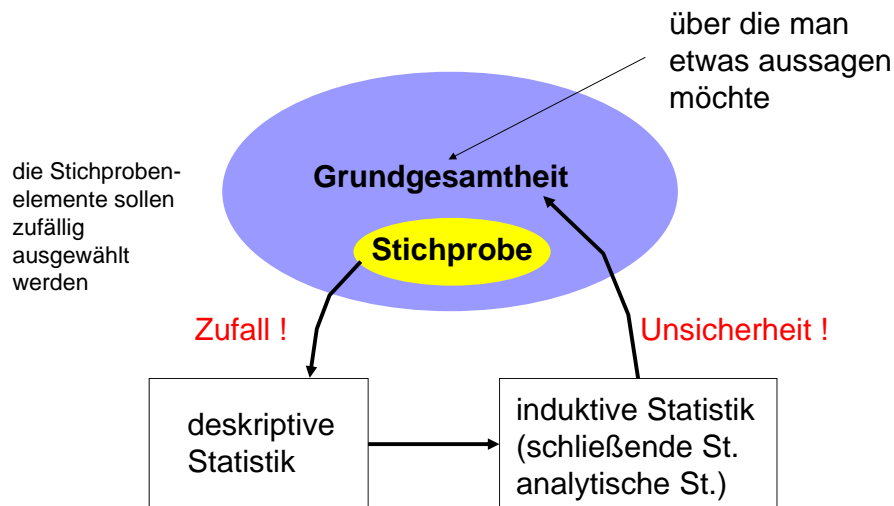
Stichprobe:

Der für die Studie ausgewählte Teil der Population.

$n = \text{endlich}$

$N \gg n$ (Umfang)

10



Die deskriptive Statistik ist die Vorstufe zur induktiven Statistik

11

Wie hoch ist die normale Pulsfrequenz (einer Population)?

Merkmal: Pulsfrequenz

zufällige Erhebung einiger

Elementen der Population: **Stichprobe**

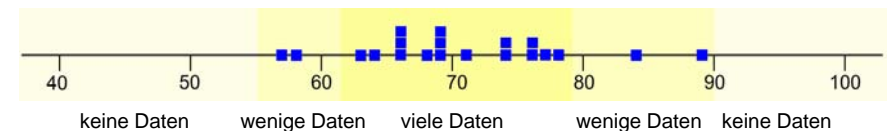
Daten der Stichprobe liegen in Form einer Urliste vor:

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!



12

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!

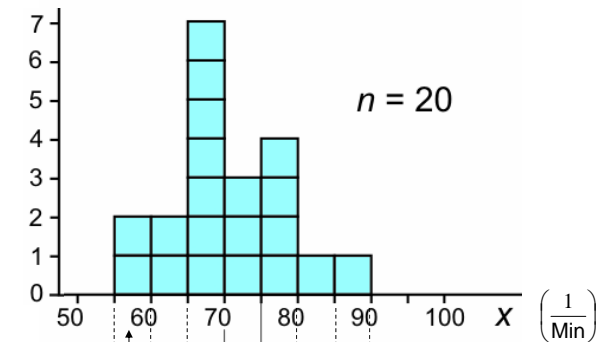
13

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:

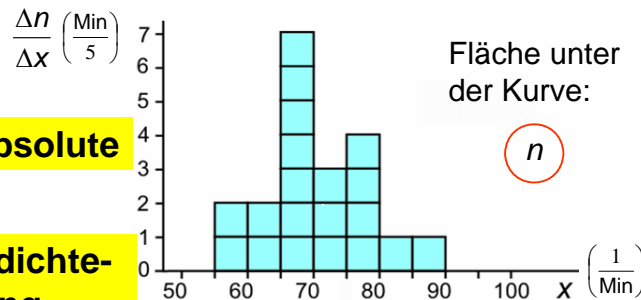
$$5 \frac{1}{\text{Min}} \cdot 2 \frac{\text{Min}}{5} = 2$$

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

Anzahl der Messdaten in der Stichprobe

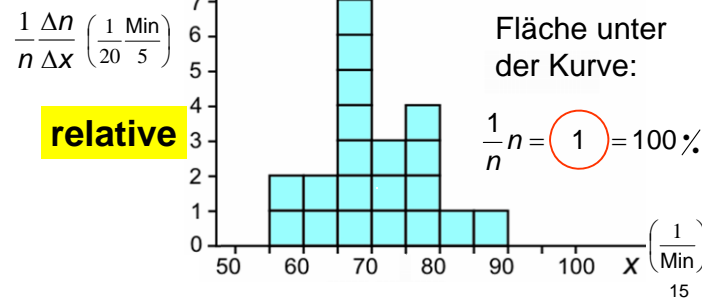
KLASSENGRENZEN	HÄUFIGKEIT
$55 \leq x_i < 60$	2
$60 \leq x_i < 65$	2
$65 \leq x_i < 70$	7
$70 \leq x_i < 75$	3
$75 \leq x_i < 80$	4
$80 \leq x_i < 85$	1
$85 \leq x_i < 90$	1
insgesamt:	$n = 20$

14



absolute

Häufigkeitsdichte-verteilung



relative

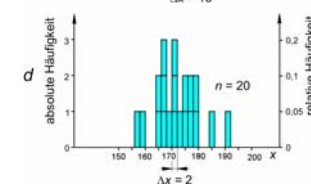
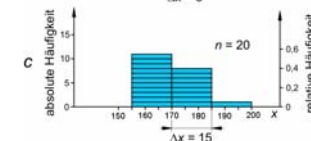
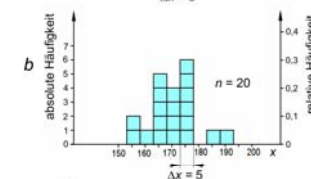
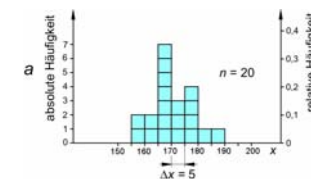
Fläche unter der Kurve:

n

Fläche unter der Kurve:

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

absolute Häufigkeits-dichte (Histogramm)

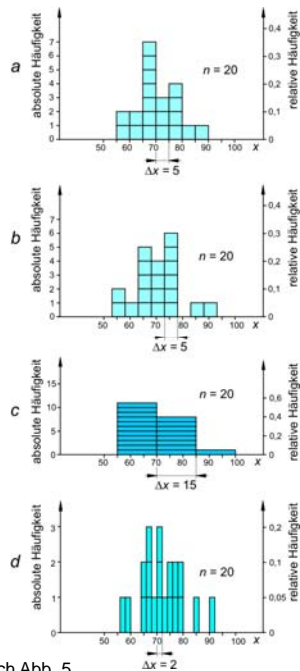


relative Häufigkeits-dichte (Histogramm)

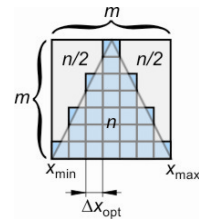
„Jedes Rechteck entspricht einem Messwert.“

Pr.Buch Abb. 5





Bestimmung der optimalen Klasseneinteilung



optimale Klassenanzahl m :

$$m^2 = 2n$$

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

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

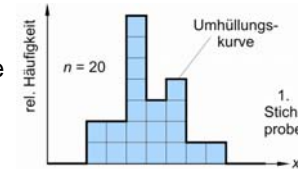
optimale Klassenbreite Δx :

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

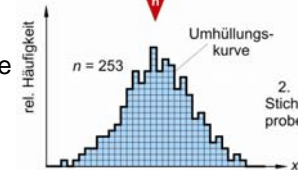
$$\Delta x = \frac{89 - 56}{6.3} = 5.2$$

Pr.Buch Abb. 5

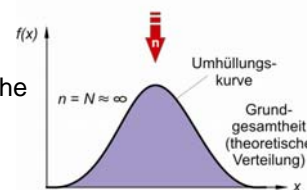
empirische Funktion



empirische Funktion



theoretische Funktion



Pr.Buch Abb. 6

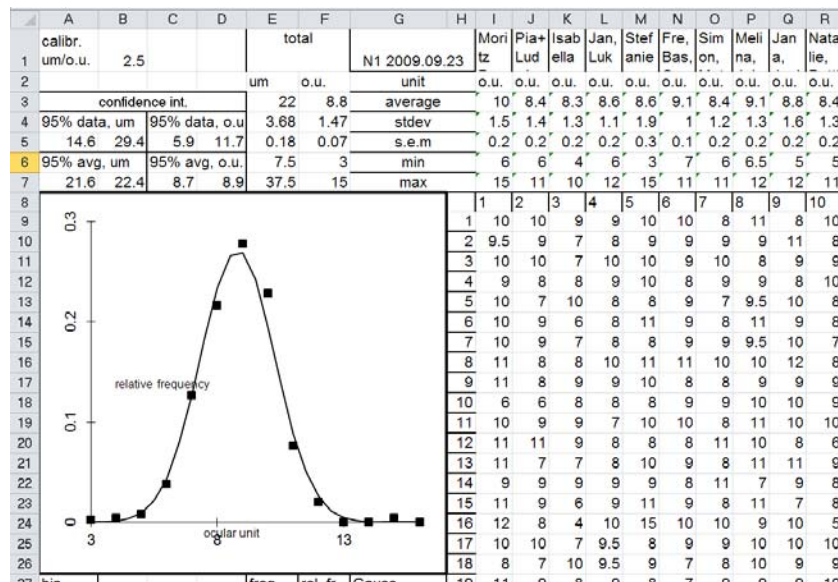


n vergrößert sich, die Klassenbreite Δx kann verkleinert werden

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

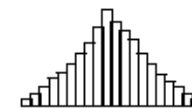
18

Beispiel: Biophysik Praktikum, Mikroskop

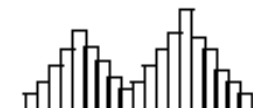


Analyse von Häufigkeitsverteilungen

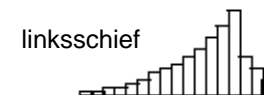
homogene symmetrische Stichprobe:



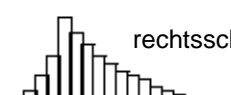
heterogene Stichprobe:



homogene nichtsymmetrische Stichproben:

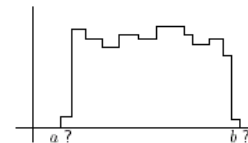


linksschief

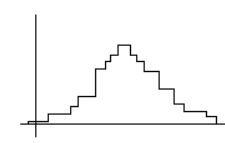


rechtsschief

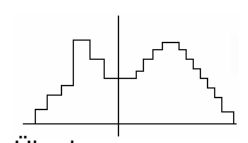
Vermutung:



Gleichverteilung?



Normalverteilung?



Überlagerung von zwei Normalverteilungen?