

Statistik

TEIL 2

Hans-Hermann Thulke
ba @ thulke-statistics.de
0172-3449934

Statistik

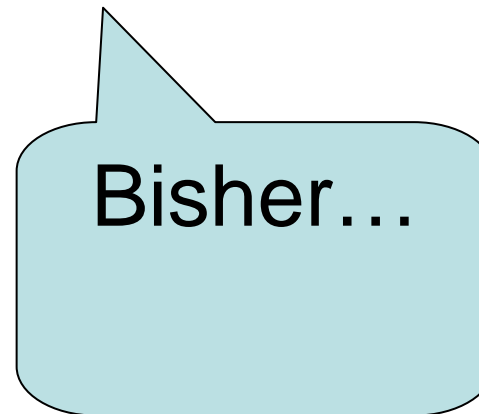
- Daten erheben, verstehen, werten
- Hypothesen prüfen
- Modellieren von Zusammenhängen

Grundgesamtheit + Stichprobe

Wahrscheinlichkeit

Datentypen, Merkmalskalen

Häufigkeits- & Punktediagramm



Bisher...

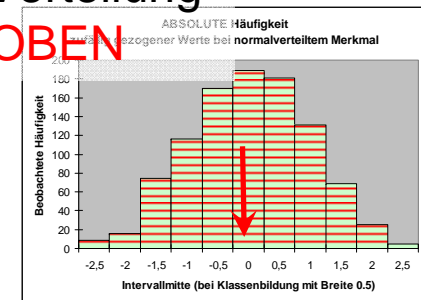
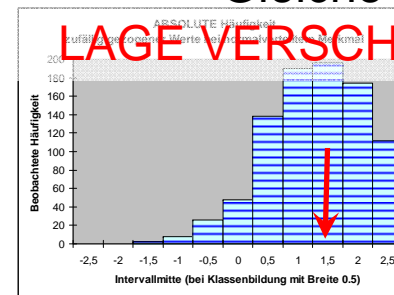
Deskriptive Statistik

d.h. Daten QUANTITATIV Beschreiben

1. Wie sind die Datenwerte verteilt? Wo liegt die Mehrzahl der Daten? **LAGEMAßE**

1. Median
2. Mittelwert
3. Modus

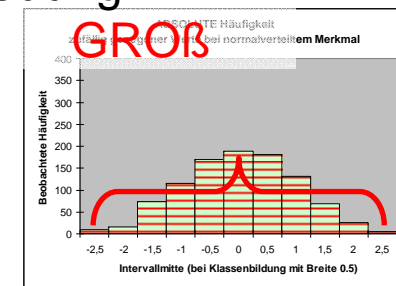
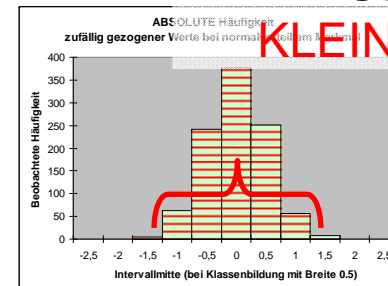
Gleiche Verteilung



2. Wie stark streuen die Daten um ihre „Lage“? Wie stark variieren die Messungen? **STREUUNGSMAßE**

1. Quantile, Spannweite
2. Varianz und Standardabweichung
3. Variationskoeffizient

Streuung



LAGEMAßE

Median

Mittiger Wert

einer Zufallsgröße / eines Merkmals
(gleich viele Werte größer bzw. kleiner)

Mitte der nach Größe geordneten Werte



Das monatliche Einkommen von sieben Anwälten:

1.000 5.000 5.000 7.000 10.000 20.000 92.000 = 7 Werte

Median: 7.000 Euro



= MEDIAN($X_1:X_n$)

Median

Mittiger Wert

(gleich viele Werte größer bzw. kleiner)

BERECHNUNG

Ist n ist die Anzahl der Beobachtungswerte x_i , dann gilt:

$$n \text{ ungerade} \Rightarrow x_{\text{Med}} = x_{\frac{n+1}{2}} \qquad n \text{ gerade} \Rightarrow x_{\text{Med}} = \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$$

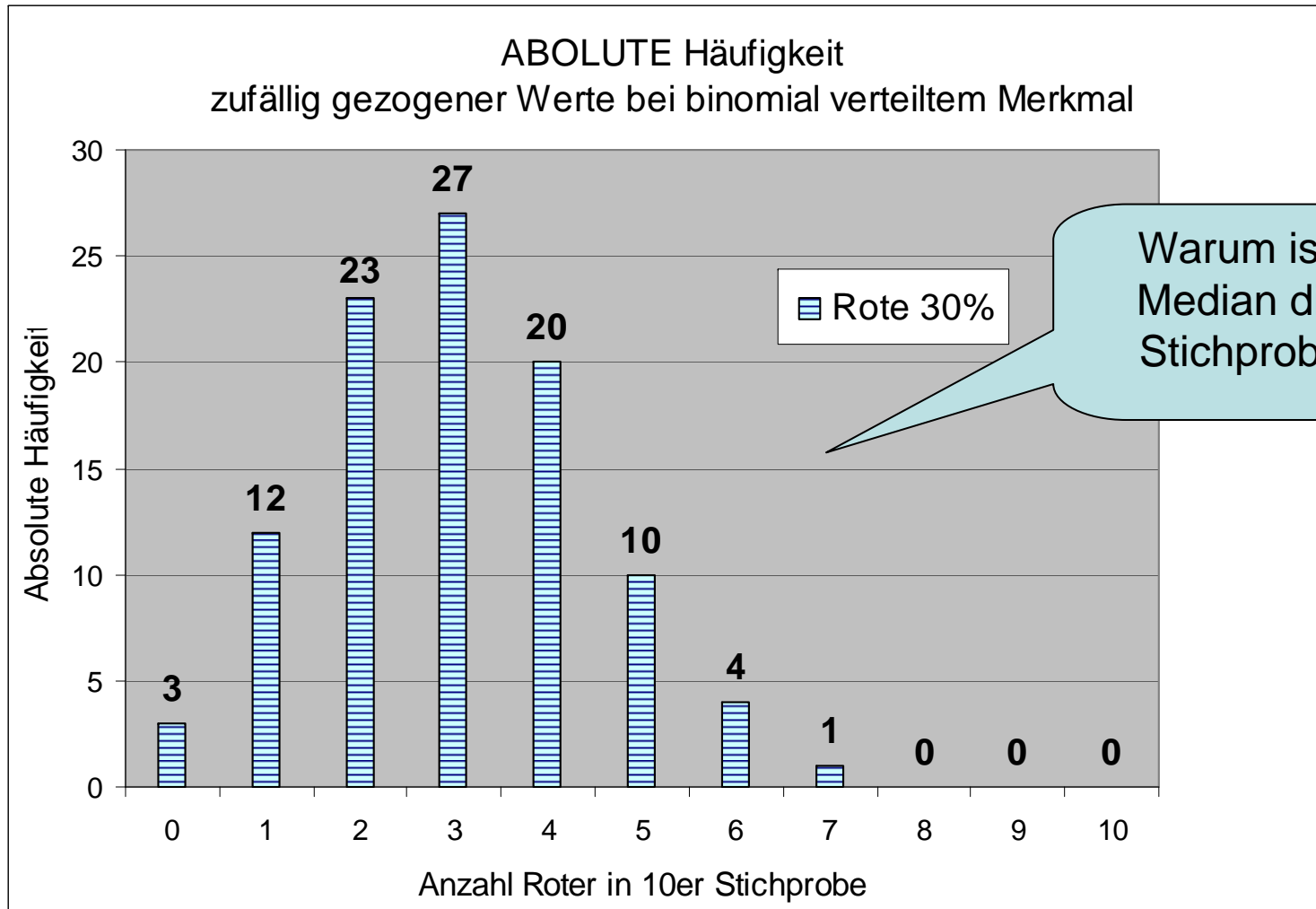
In Verallgemeinerung des empirischen Medians heißt für eine Zahl p mit $0 < p < 1$

$$x_p := \begin{cases} x_{([n \cdot p + 1])} & , \text{ falls } n \cdot p \notin \mathbb{N}, \\ \frac{1}{2} \cdot (x_{(n \cdot p)} + x_{(n \cdot p + 1)}) & , \text{ falls } n \cdot p \in \mathbb{N}. \end{cases}$$

Für $p=0.5$ oder 50%
ergibt sich der Median!!!

das (**empirische**) **p -Quantil** (engl.: *sample p -quantile*) von x_1, \dots, x_n . Dabei bezeichnet allgemein der Ausdruck $[y] := \max\{k \in \mathbb{Z} : k \leq y\}$ die größte ganze Zahl, welche kleiner oder gleich einer reellen Zahl y ist, also z.B. $[1.2] = 1$, $[-0.3] = -1$, $[5] = 5$.

Beispiel



Mittelwert

„laut Statistik haben ein Millionär und ein armer Kerl jeder eine halbe Million.“
(Franklin D. Roosevelt, 32. Präsident USA)

Durchschnittliche Wert

einer Zufallsgröße / eines Merkmals

Engl. Mean; Summe geteilt durch Anzahl

Das monatliche Einkommen von sieben Anwälten:

$$1.000 + 5.000 + 5.000 + 7.000 + 10.000 + 20.000 + 92.000 = 140.000$$

Median: 7.000 Euro

Mittelwert: $140.000 / 7 = 20.000$ Euro

Mittelwert

Durchschnittliche Wert

Summe geteilt durch Anzahl

BERECHNUNG

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

n : Anzahl der Beobachtungswerte x_i

x_i : i -ter Beobachtungswert $i \in \mathbb{N}$



= MITTELWERT(X1:Xn)

= SUMME(X1:Xn) / ANZAHL(X1:Xn)

Natürliche Zahlen

(Wie sind die Symbole für Ganze Zahlen oder Reelle Zahlen ??)

Modus

Wert, der häufiger vorkommt als jeder andere

einer Zufallsgröße / eines Merkmals

Auch Modalwert (engl: mode)

(Häufigste Wert = höchster Balken in Häufigkeitsdiagramm)

Das monatliche Einkommen von sieben Anwälten:

1.000 5.000 5.000 7.000 10.000 20.000 92.000

Median: 7.000 Euro

Mittelwert: $140.000 / 7 = 20.000$ Euro

Modus: 5.000 Euro

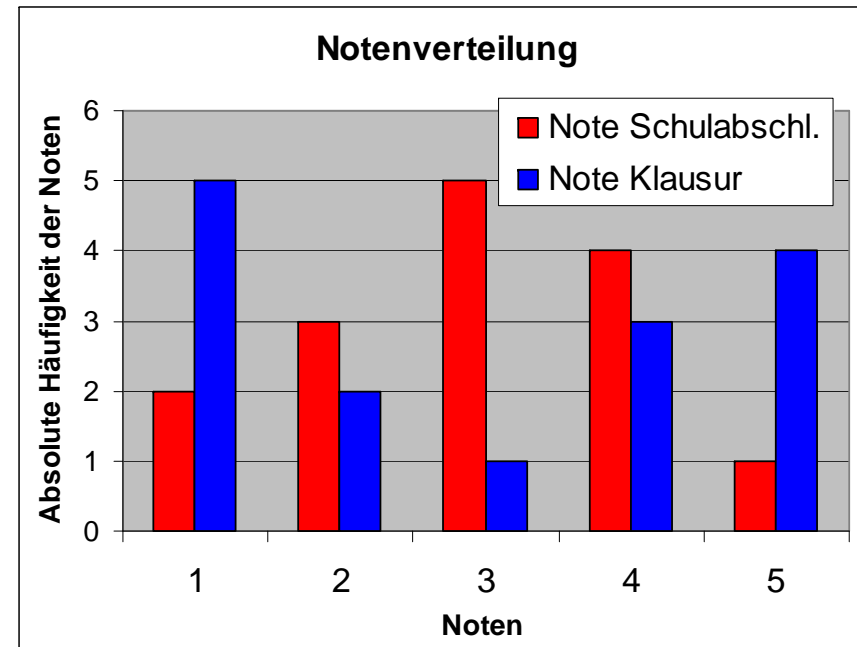


= MODALWERT($X_1:X_n$)

Häufigkeitstabelle

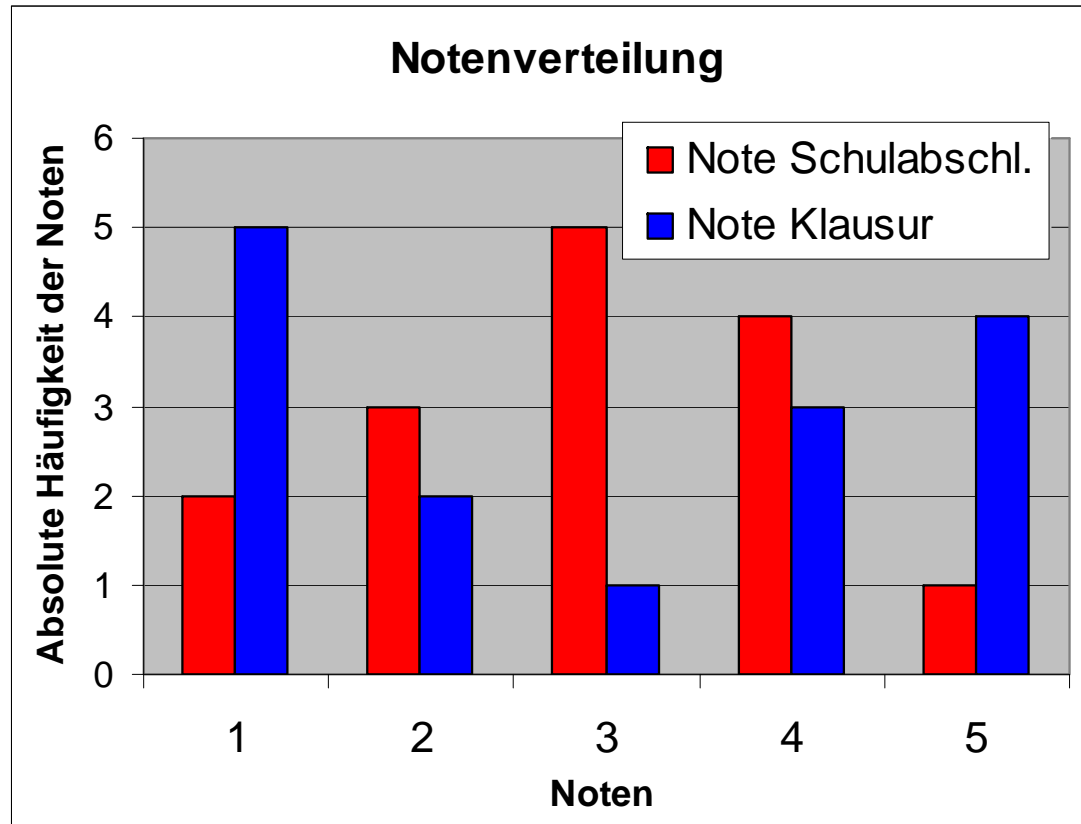
Student	Note Schulabschl.	Note Klausur
1	1	1
2	2	1
3	3	5
4	2	5
5	1	4
6	5	2
7	3	3
8	4	4
9	4	1
10	2	5
11	4	1
12	4	2
13	3	1
14	3	5
15	3	4

Beispiel



Lagemaß	Note Schulabschl.	Note Klausur
Median	3	3
Mittelwert	2,93	2,93
Modus	3	1

Beispiel



2,93 Mittelwert 2,93

Unterschiedlich für beide Datensätze ist die STREUUNG um den Mittelwert:

ROT mehr Werte in der Nähe des Mittelwerts (3 Werte weichen um 2 Noten ab)

BLAU mehr Werte fern vom Mittelwert (9 Werte weichen um 2 Noten ab)

STREUUNGSMAßE

Spannweite

$$„X_{\max} - X_{\min}“$$

Größter minus kleinster Wert

einer Zufallsgröße / eines Merkmals

Engl. Range

Minimum

Maximum

Das monatliche Einkommen von sieben Anwälten:

1.000 5.000 5.000 7.000 10.000 20.000 92.000

Median: 7.000 Euro

Mittelwert: $140.000 / 7 = 20.000$ Euro

Modus: 5.000 Euro

Spannweite: 91.000 Euro



$$= \text{MAX}(X_1:X_n) - \text{MIN}(X_1:X_n)$$

Interquartilsabstand

$$„X_{75\%} - X_{25\%}“$$

($X_{0,75} - X_{0,25}$ siehe Medianfolie)

Differenz zwischen 3. und 1. Quartil

einer Zufallsgröße / eines Merkmals

Engl. Inter quartils range

$$p=0.25; n=7; np=1,75$$

$$X_{0,25} = X_{[np + 1]} = X_{[2,75]} = X_2$$

$$p=0.75; n=7; np=5,25$$

$$X_{0,75} = X_{[np + 1]} = X_{[6,25]} = X_6$$

Das monatliche Einkommen von sieben Anwälten:

1.000 5.000 5.000 7.000 10.000 20.000 92.000

Median: 7.000 Euro; Mittelwert: $140.000 / 7 = 20.000$ Euro; Modus: 5.000 Euro

Spannweite: 91.000 Euro; **Quartilsabstand: 15.000** aus $20.000 - 5.000$ (siehe Medianfolie!!!)



$$= \text{QUARTILE}(X_1:X_n; 3) - \text{QUARTILE}(X_1:X_n; 1)$$

Varianz

Gemittelte quadrierte Abweichungen vom Mittelwert

einer Zufallsgröße / eines Merkmals

Engl. variance

Das monatliche Einkommen von sieben Anwälten:

$$1.000 + 5.000 + 5.000 + 7.000 + 10.000 + 20.000 + 92.000 = 140.000$$

Median: 7.000 Euro

Mittelwert: $140.000 / 7 = 20.000$ Euro

Modus: 5.000 Euro

Spannweite: 91.000 Euro; Quartilsabstand: 15.000

Varianz (SP): 1.044.000.000 Euro² / Standardabweichung: 32.311 Euro



Varianz

Unterschied?

Gemittelte quadrierte Abweichungen vom Mittelwert

BERECHNUNG (Stichprobe!!!)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right)$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

Ausnutzen für Handrechnung:
VERSCHIEBESATZ

= VARIANZ(X₁:X_n)

- :: berechne je Wert X_i die quadrierte Abweichung vom Mittelwert
= (X_i - MITTELWERT(X₁:X_n)) * (X_i - MITTELWERT(X₁:X_n))
- :: dann ist die Varianz (SP) zu berechnen als:
= SUMME(X-AQ₁:X-AQ_n) / (ANZAHL(X₁:X_n) -1)

Standardabweichung

Wurzel aus der Varianz

Engl: standard deviation

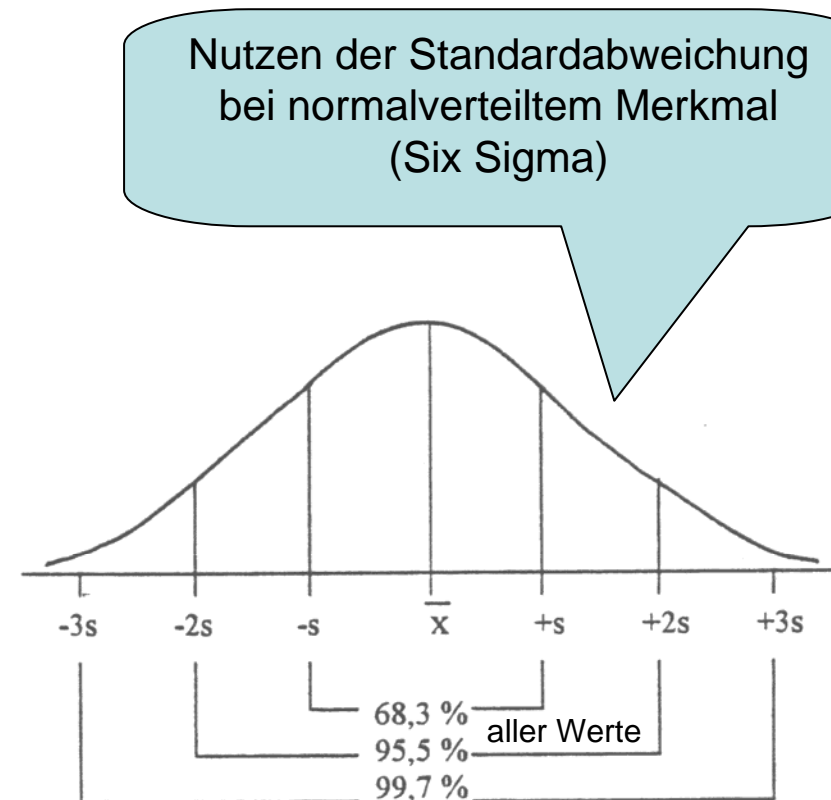
BERECHNUNG (Stichprobe!!!)

$$s = \text{Wurzel} (s^2) = \sqrt{\frac{\sum (x - \bar{x})^2}{(n - 1)}}$$



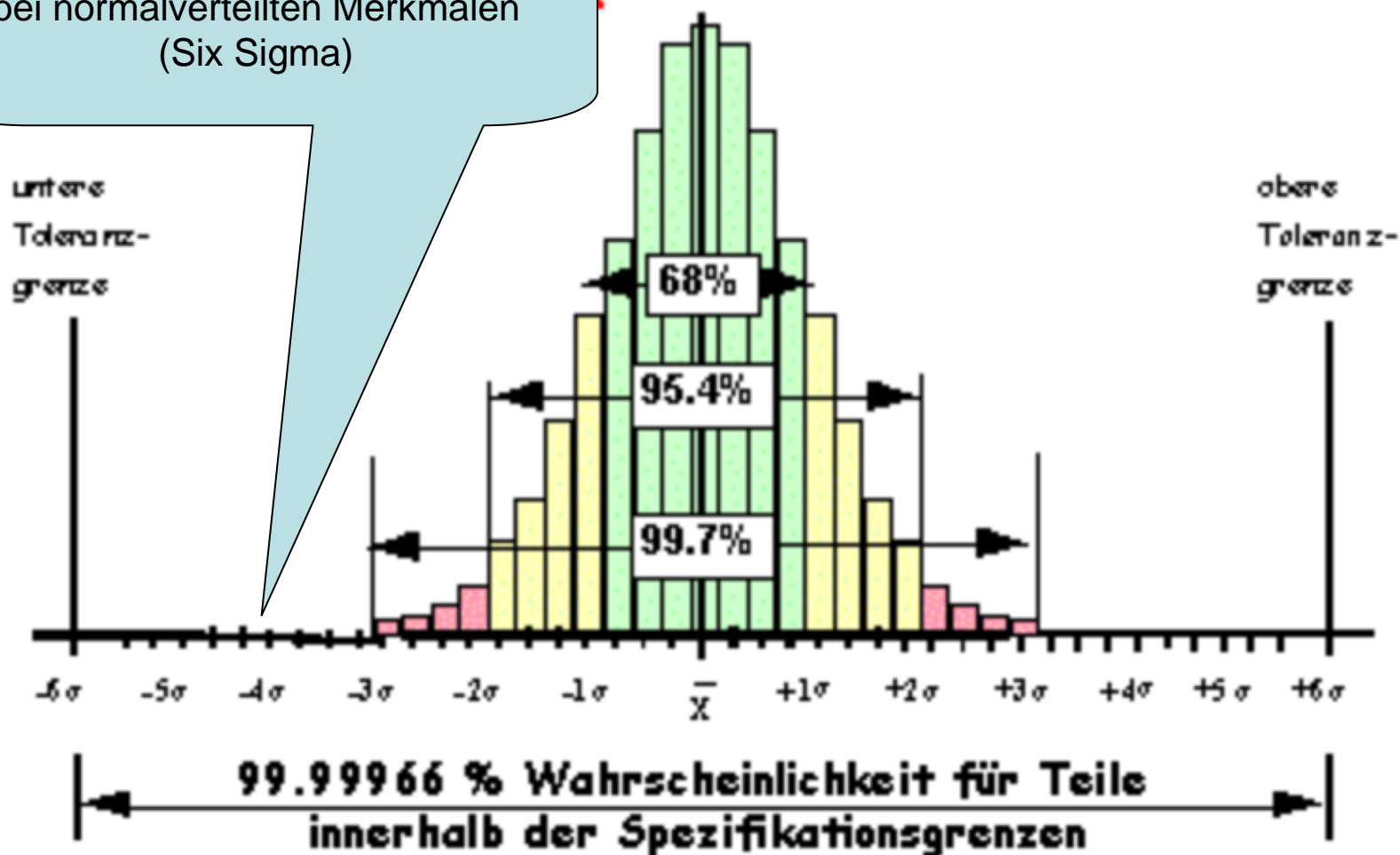
= STABW($X_1:X_n$)

= WURZEL (VARIANZ($X_1:X_n$))



Standardabweichung

Nutzen der Standardabweichung
bei normalverteilten Merkmalen
(Six Sigma)



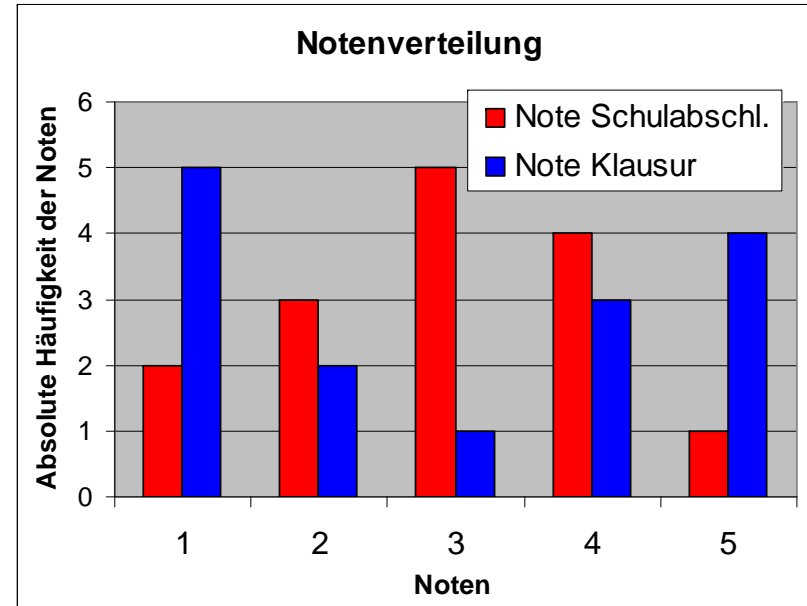
Student	Note Schulabschl.	Note Klausur
1	1	1
2	2	1
3	3	5
4	2	5
5	1	4
6	5	2
7	3	3
8	4	4
9	4	1
10	2	5
11	4	1
12	4	2
13	3	1
14	3	5
15	3	4

Beispiel

Unterschiedlich für beide Datensätze ist die STREUUNG um den Mittelwert:

ROT mehr Werte in der Nähe des Mittelwerts (3 Werte weichen um 2 Noten ab)

BLAU mehr Werte fern vom Mittelwert (9 Werte weichen um 2 Noten ab)



Mittelwert war **2,93** = **2,93**

Varianz **1,35** < **2,92**

Standardabweichung **1,16** < **1,71**

STREUUNG um den Mittelwert:

ROT mehr Werte in der Nähe des Mittelwerts (3 Werte weichen um 2 Noten ab) – **Streuung klein**

BLAU mehr Werte fern vom Mittelwert (9 Werte weichen um 2 Noten ab) - **Streuung größer**

Variationskoeffizient

„Relative Streuung“

Verhältnis Standardabweichung zu Mittelwert

einer Zufallsgröße / eines Merkmals

Engl. Variation coefficient

$$V = \frac{s}{\bar{x}} \cdot 100 \%$$

Das monatliche Einkommen von sieben Anwälten:

$$1.000 + 5.000 + 5.000 + 7.000 + 10.000 + 20.000 + 92.000 = 140.000$$

Median: 7.000 Euro; Mittelwert: 20.000 Euro; Modus: 5.000 Euro

Spannweite: 91.000 Euro; Quartilsabstand: 15.000

Varianz (SP): 1.044.000.000 Euro² bzw. Standardabweichung: 32.311 Euro

Variationskoeffizient: 1,62

Grundgesamtheit + Stichprobe
Wahrscheinlichkeit
Datentypen, Merkmalskalen
Häufigkeits- & Punktediagramm
Lagemaße & Streuungsmaße

Mögliche Aufgabe

GEGEBEN: 4 Datensätze á 50 Werte als Häufigkeitstabelle

	X	0	1	2	3	4	5	6	7	8	9
Häufigkeit	V1	1	2	5	7	10	10	7	5	2	1
V2	1	5	10	6	3	4	6	7	7	1	
V3	0	3	5	7	10	10	7	5	3	0	
V4	1	5	10	5	4	4	5	10	5	1	

GESUCHT: Ergänzen Sie die Deskriptoren zu den Datensätze V1-V4.

Häufigkeitsdiagramm	Varianz	Standardabweichung	Mittlere Abweichung	Median
	V.. 3,8 V..	2,5 V..	1,5 V..	4,5 V..
	V.. ?? V2	2,0 V..	2,3 V..	?? V2
	V.. 3,5 V..	1,9 V..	2,3 V..	4,5 V..
	V.. 6,4 V..	?? V2	1,6 V..	?? V4

Veranschaulichung

Lage- und Streunungsmaße

1. <http://www.tipptreffer.de/lotto/lottosummen.htm>
2. NOCHMAL GANZ IN RUHE:
<http://www.mathe-online.at/clips/mwstdabw/index.html>
(siehe auch das „Zusatzmaterial“)