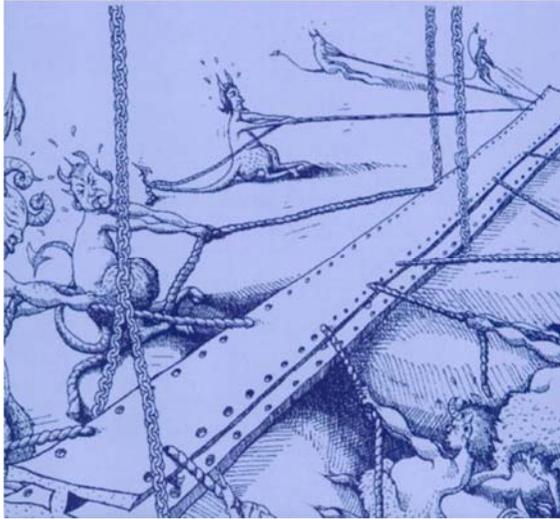


Regression und Korrelation

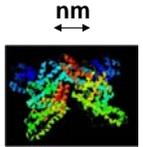


regression:
Zurückführung,
Rückschreiten

correlation:
Wechselbeziehung

KAD 2016.02.16

Praktische Annäherung (Beispiel1)



1 St. HSA Molekül

wieviele Eiweissmoleküle sind in dem Blutplasma?
(Stück, mol, g, ...)

wie gross ist die Eiweisskonzentration
des Blutplasmas? (St/L, mol/L, g/L)

bei Patienten in Nephrose (schwere Nierenkrankheit) nimmt der Wert stark ab

direkte Methode:

Bestimmung der Anzahl der Moleküle in einem Volumen(?)

indirekte Methode :

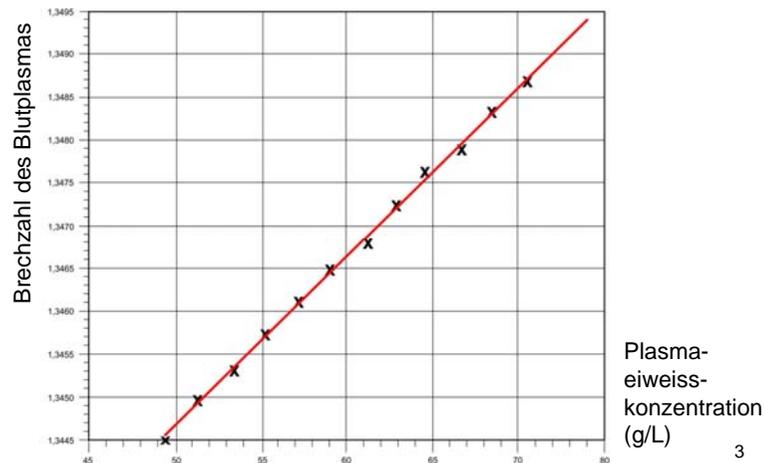
mit Hilfe einer (einfach) messbaren physikalischen Grösse, die
steht in streng monoton wachsendem Zusammenhang zu der
unbekannten Grösse

(die solche einfachste Funktion ist ...)

2

Bemerkung:

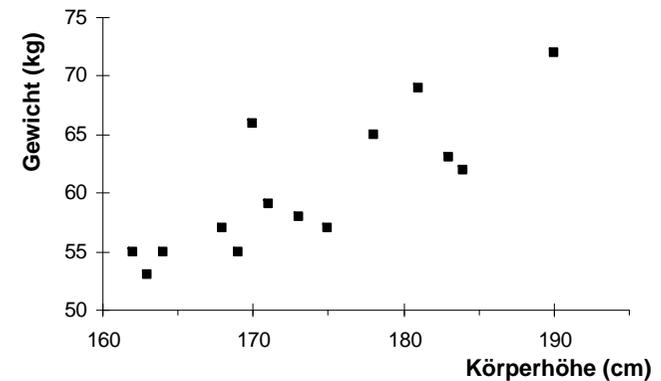
das Licht breitet sich in Blutplasma langsamer, wenn die
Plasmaeiweisskonzentration grösser ist, d.h. das Licht hat
grössere Brechzahl (deterministischer Zusammenhang, Messfehler)



3

(Beispiel2)

Daten aus einer Studentengruppe E2
(Sept. 1994) (zusammengehörige
Wertepaare)



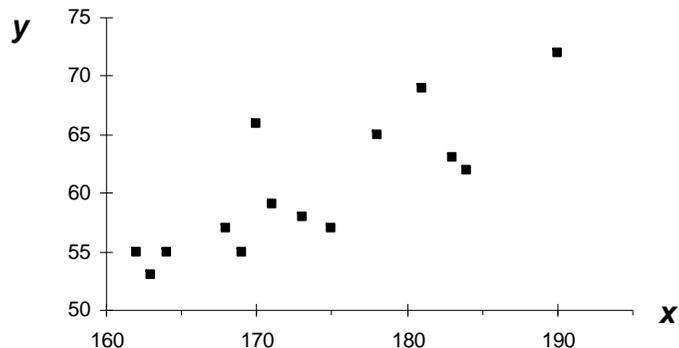
cm	kg
162	55
163	53
164	55
168	57
169	55
170	66
171	59
173	58
175	57
178	65
181	69
183	63
184	62
190	72

was für eine Tendenz kann man bemerken?

4

Die Korrelationsrechnung beschäftigt sich mit dem symmetrischen Zusammenhang zweier Zufallsgrößen

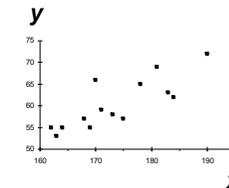
positive Korrelation: je mehr, desto mehr
negative Korrelation: je mehr, desto weniger



hier: positive Korrelation

Regressionsannäherung

Sucht man einen Funktionszusammenhang zwischen einer (oder mehreren) unabhängigen Variable (x) und einer abhängigen Variable (y)



Voraussetzungen: x und y numerische und stetige Merkmale, y Zufallsgröße (ihre Größe wird nicht nur von der unabhängigen Variable, sondern durch den Zufall beeinflusst)

(a: Steigung, b: Achsenabschnitt)

Regressionsmodell fixiert den Typ der Funktion:

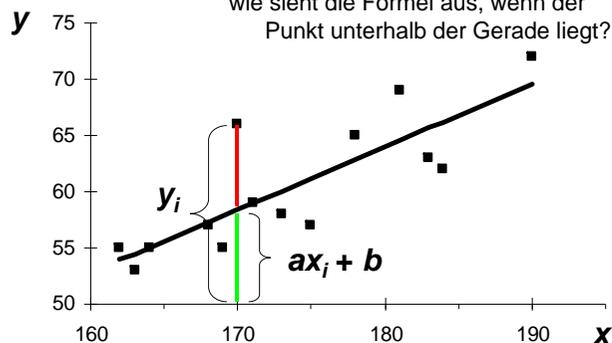
- lineare F. $y = (ax + b) + h$
- polinomiale F. $y = a + b_1x + b_2x^2 + \dots + b_nx^n + h$
- exponentiale F. $y = ab^x h$
- Potenzfunktion $y = ax^b h$

und wie wirkt der Zufall auf die abhängige Variable
additiver Fehler (+ h) oder multiplikativer Fehler (·h)

Das einfachste Regressionsmodell: lineare Regression

lineare Funktion: $y = (ax + b) + h$

$h_i = y_i - (ax_i + b)$ wenn der Punkt (x_i, y_i) oberhalb der Gerade liegt
wie sieht die Formel aus, wenn der Punkt unterhalb der Gerade liegt?



Beste Gerade: Summe der Fehlerquadrate ist minimal (Methode der kleinsten Quadraten)

	x_i	y_i
1	162	55
2	163	53
3	164	55
4	168	57
5	169	55
6	170	66
7	171	59
8	173	58
9	175	57
10	178	65
11	181	69
12	183	63
13	184	62
14	190	72

„Die beste“ Steigung:

$(y = ax + b)$

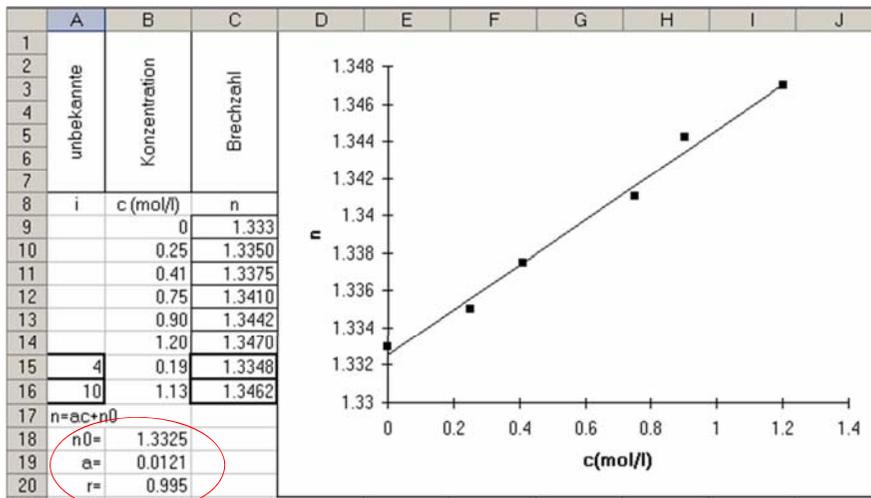
$$a^* = \frac{Q_{xy}}{Q_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{oder} \quad a^* = \frac{s_{xy}^2}{s_x^2}$$

„Der beste“ Achsenabschnitt:

$$b^* = \bar{y} - a^* \cdot \bar{x} = \frac{\sum_{i=1}^n y_i}{n} - a^* \frac{\sum_{i=1}^n x_i}{n}$$

wo $s_{xy}^2 = \frac{Q_{xy}}{n-1}$: Kovarianz

Beispiel: Refraktometrie



9

Wie gut passen die Messpunkte an die Regressionsgerade?

Korrelationsrechnung beschreibt die lineare Beziehung zwischen zwei oder mehr statistischen Variablen

es beschreibt die Stärke der Korrelation
es gibt starke und schwache Korrelation

Korrelationskoeffizient (Pearson)

$$r = \frac{Q_{xy}}{\sqrt{Q_{xx} \cdot Q_{yy}}} = \frac{S_{xy}^2}{S_x S_y}$$

der Zähler ist gleich dem Zähler der Steigung der Regressionsgerade (der Nenner ist im beiden Fall positiv)

$$a^* = \frac{Q_{xy}}{Q_{xx}}$$



positive Steigung: r ist positive Zahl
negative Steigung: r ist negative Zahl

$$-1 \leq r \leq 1$$

10

weitere Bemerkungen:

$$-1 \leq r \leq 1$$

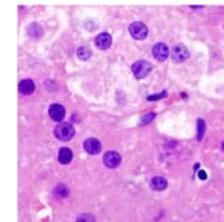
Korrelationskoeffizient (Pearson)

$$0 \leq r^2 \leq 1$$

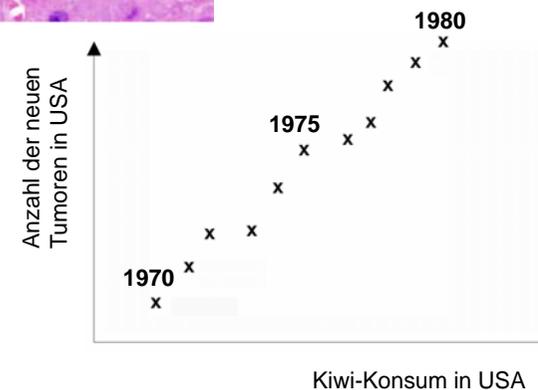
Bestimmtheitsmass (coefficient of determination)

Die Korrelation beschreibt nicht unbedingt eine Ursache-Wirkungs-Beziehung in die eine oder andere Richtung.

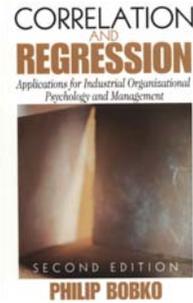
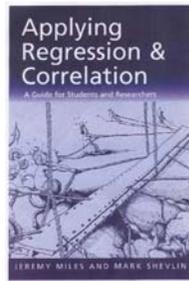
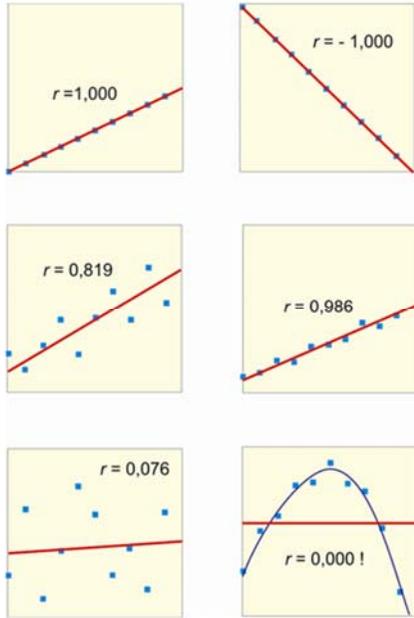
11



Korreliert heisst nicht notwendigerweise kausal verknüpft(!)

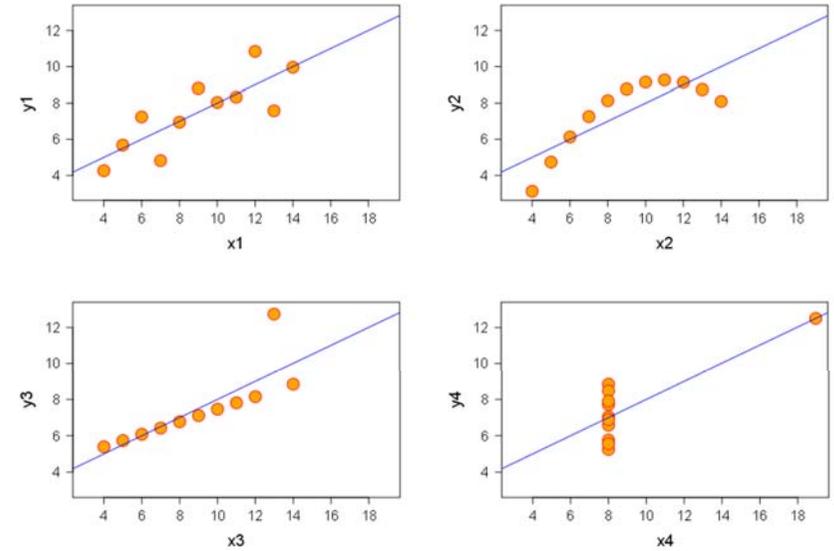


Beispiele:



Pr.Buch Abb. 15

Extrembeispiel: $r=0.816, y = 3 + 0.5x$ (Anscombe's quartet)

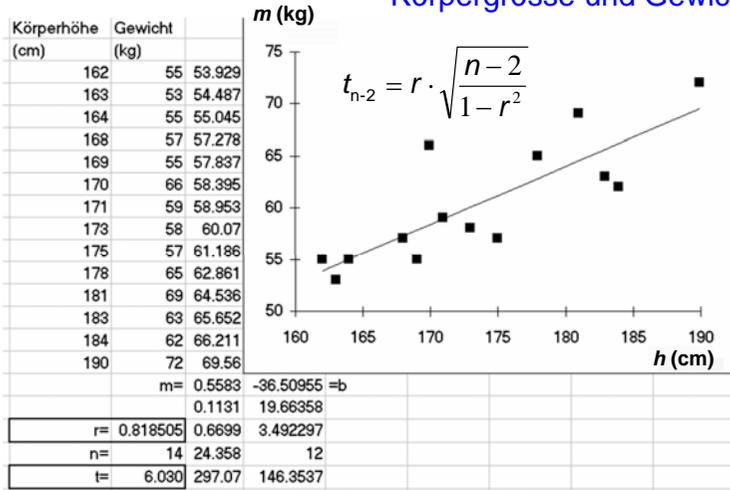


http://en.wikipedia.org/wiki/Anscombe%27s_quartet

14

t-Test zur Korrelationsanalyse

Gibt es eine Beziehung zw. der Körpergröße und Gewicht?



$|t| = 6.030 > t_{12, \text{krit}(0,05)} = 2.179 \Rightarrow H_0 \text{ ist falsch (} p < 0.05 \text{)}$
 $|t| = 6.030 > t_{12, \text{krit}(0,01)} = 3.055 \Rightarrow H_0 \text{ ist falsch (} p < 0.01 \text{)}$

15

Weiteres Beispiel: Leistung der Röntgen-Röhre

