

~~Anmeldung bis 15.08.~~

~~Anmeldung bis 30.08.~~

~~Rockladen / Abgabe 31.08.~~

## Mehrdimensionales Datenmaterial

- Zwei-dimensionales Datenmaterial

verbundene Stichprobe vom Umfang  $n$

$(x_1, y_1); (x_2, y_2); \dots; (x_n, y_n)$

$X$ : ... habe die Ausprägungen

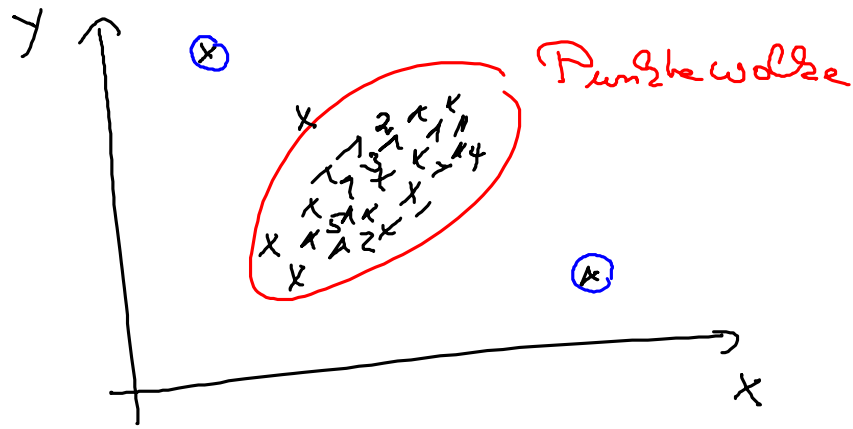
$a_1; a_2; \dots; a_k$

$Y$ : ... habe die Ausprägungen

$b_1; b_2; \dots; b_l$

# Darstellungsformen

a) wenn  $X$  und  $Y$  kardinal skaliert  
Streudiagramm



⊗ Ausreißer

## Aufgaben

① ist der lineare Ansatz okay?

② genügend Datenpunkte?  
mind. 8

③ auffällige Datenpunkte?  
Ausreißer

b) bei anderen Skalierungspunkten  
Kontingenztafel, Kontingenz-  
tabelle

X \ Y	$b_1$	$b_2$	...	$b_j$	...	$b_l$	Y ist egal
$a_1$	$h_{11}^o$	$h_{12}^o$	...	$h_{1j}^o$	...	$h_{1l}^o$	$h_{1\cdot}^o$
$a_2$	...	...	...	...	...	...	$h_{2\cdot}^o$
$\vdots$							$\vdots$
$a_i$	...	...	...	$h_{ij}^o$	...	...	$h_{i\cdot}^o$
$\vdots$							$\vdots$
$a_l$	$h_{l1}^o$	$h_{l2}^o$	...	$h_{lj}^o$	...	$h_{ll}^o$	$h_{l\cdot}^o$
	$h_{\cdot 1}$	$h_{\cdot 2}$	...	$h_{\cdot j}$	...	$h_{\cdot l}$	n

gemeinsame Käuf-  
zeilen  
flächig

RANDHÄUFIGK.

X ist egal

Y ist egal

$n$ : Stichprobenumfang

$h_{ij}^o$ : beobachtete gemeinsame Käuf-  
zeilen  
Anzahl der Fälle, bei denen  
X nach  $a_i$  und Y nach  $b_j$   
realisiert sind

$b_l$ : Anzahl unterschiedl. Ausprägungen  
bei X

$l$ : Anzahl unterschiedl. Ausprägungen

bei  $\gamma$

$h_{i\cdot}$ : eindimensionale H-Verteilung von X

$h_{\cdot j}$ : " " " " "  $Y$

mit von Gültigkeit - Anfangs

X: Wohnverhalten der Väter A, B, C

Y: " " Söhne A, B, C

Aufg. 7

X \ Y	A	B	C	
A	10	20	20	50
B	50	15 <sub>a</sub>	5 <sub>b</sub>	120
C	20	5 <sub>c</sub>	5 <sub>d</sub>	30
	80	40	80	200

1. Spiegel punkt

Randhäufigkeiten (RH)

$$2: 1: 2 \Rightarrow 2 + 1 + 2 = 5 \text{ Teil}$$

$$200: 5 = 40 / \text{Teil}$$

## 2. Spiegel punkt

gemeinsame H.

5% von 200 sind 10

$$n_{AA} = 10$$

## 3. Spiegel punkt

400 Stimmen

40% von 400  
sind 160

RH

$$n_{B\cdot} + n_{\cdot B} = 160$$

B-Stimmen

$$n_{B\cdot} + 40 = 160 \Rightarrow \underline{\underline{n_{B\cdot} = 120}}$$

## 4. Spiegel punkt RH

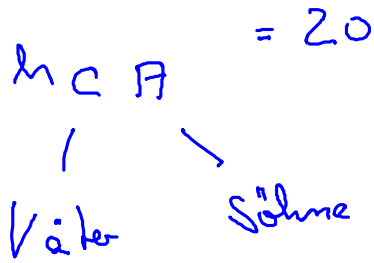
15% von 200 Stimmen sind 30

## 5. Spiegel punkt gemeinsame H.

$$n_{AB} = 20$$

Väter      Söhne

6. Spiegel punkt gemeinsame H.



7. Spiegel punkt gemeinsame H.

$\Rightarrow$  LGS

$$\text{I} \quad a + b = 70$$

$$\text{II} \quad c + d = 10$$

$$\text{III} \quad d + c = 20$$

$$\text{IV} \quad b + d = 60$$

$$\text{V} \quad a + d = 20 \quad \text{7. Spiegel punkt}$$

$$\text{aus IV} \Rightarrow b = 60 - d$$

$$\text{aus V} \Rightarrow a = 20 - d$$

$$\text{in I} \Rightarrow 20 - d + 60 - d = 70$$

$$80 - 2d = 70 \quad | -80$$

$$-2d = -10 \quad | :(-2)$$

$$\underline{\underline{d = 5}}$$

Nicht von Wichtigkeit - Ende

## Abhängigkeitsanalyse

1.) lineare Einfachregression

Wichtig!!!

2.) Modelle der Korrelationsrechnung

→ Ziel:

Feststellen der Stärke (perfekt, stark, mäßig, schwach, kein) kein festes MUSS

konkret

1)  $x \uparrow \Leftrightarrow y \uparrow$   
 $x \downarrow \Leftrightarrow y \downarrow$

und der Richtung  
(positiv  $\rightarrow$  gleichgerichtet, 1)  
negativ  $\rightarrow$  entgegengesetzt) 2

2)  $x \uparrow \Leftrightarrow y \downarrow$   
 $x \downarrow \Leftrightarrow y \uparrow$

des linearen Zusammenhangs  
zweier Größen X und Y.

Fall 1. X und Y sind ordinal  
skaliert (Bravais-Pearson)

=> Korrelationskoeffizient  $r$

nach Bravais-Pearson

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}}$$

$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_i^2$
8	5	40	64	25
2	7	14	4	49

$n$ : Stichprobenumfang  
Anzahl der Datenpaare

$\bar{x}, \bar{y}$ : arithmetische Mittel

Ans w.

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$\bar{y} = \frac{1}{n} \sum y_i$$

Ablauf: ① Art der Datenerhebung

$(x_i; y_i)$  **paarw.**

$X$  und  $Y$  werden gleichzeitig erhoben und  
variieren gemeinsam. **in Paaren**

Aufgabe 90 - Teil 1

$X$ : Alter (in Jahren)

$Y$ : Gewicht (in kg)



1)  $(x_i, y_i)$  Formel ③

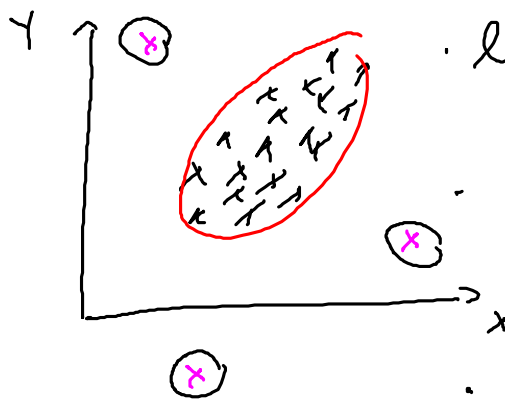
5 Kinder wurden zugleich nach Alter und Gewicht erfasst. in 2. Schritt

② Streu (unp) Diagramm

Bewertungskriterien

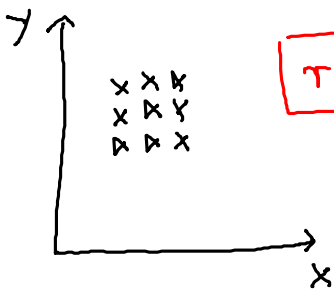
- linearer Ansatz okay?
- genügend Datenpunkte? mind. 8
- auffällige Datenpunkte? Ausreißer

Bsp. 1



- linearer Ansatz okay
- genügend Punkte
- keine Ausreißer

Bsp. 2a



$r = 0$

linearer Ansatz nicht okay

Bsp. 2b

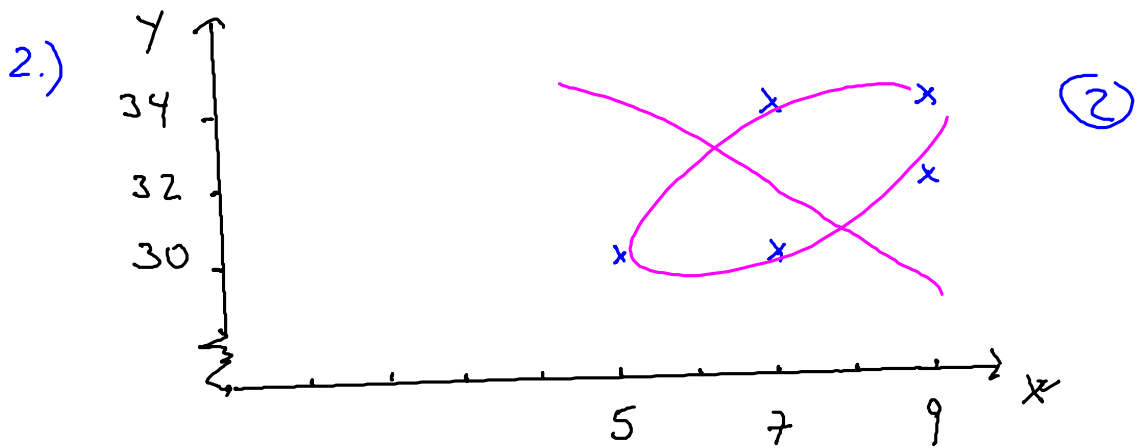


$r = 0$

kein Zusammen-  
hang zwischen  
X und Y

nicht-linearer  
Zusammenhang  
zwischen X und Y

## Aufgabe 90 - Teil 1



- durch ein lineares Ansatz  
auch fälschlicherweise möglich
- nur wenige Datenpunkte

③ Berechnung von  $r$  (Bravais-Pearson)  
und Interpretation

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2) \cdot (\sum y_i^2 - n \bar{y}^2)}}$$

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

Stichprobe - negativ

positiv

Korrelation

1.  $r = 1$

In der Stichprobe liegt ein perfekter, positiver, linearer Zusammenhang zwischen  $X$  und  $Y$

oder

!

Alle Punkte liegen auf einer geraden Gerade mit positiver Steigung

2.  $0,7 < r < 1$

...  
starker, positiver, linearer  
...

3.  $0,3 \leq r \leq 0,7$

...  
mäßiger, positiver, linearer  
...

4.  $0 < r < 0,3$

...  
schwacher, positiver, linearer  
...

5.  $r = 0$

In der Stichprobe liegt kein linearer Zusammenhang zwischen  $X$  und  $Y$

www.

→ Seite oben  
Bsp. 20 und 26.

3. Aufgabe 90 - Teil 1

Standardtabelle

$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_i^2$
5	30	150	25	900
9	34			
7	34			
9	32			
7	30			
<u><math>\sum x_i</math></u> <sup>1</sup>	<u><math>\sum y_i</math></u> <sup>2</sup>	<u><math>\sum x_i y_i</math></u> <sup>3</sup>	<u><math>\sum x_i^2</math></u> <sup>4</sup>	<u><math>\sum y_i^2</math></u> <sup>5</sup>
37	160	1192	285	5.136

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}}$$

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$\bar{y} = \frac{1}{n} \sum y_i$$

$$n = 5$$

$$\bar{x} = \frac{1}{5} \cdot 37 = \underline{\underline{7,4}}$$

$$\bar{y} = \frac{1}{5} \cdot 160 = \underline{\underline{32}}$$

$$r = \frac{1192 - 5 \cdot 7,4 \cdot 32}{\sqrt{(285 - 5 \cdot 7,4^2)(5 \cdot 136 - 5 \cdot 32^2)}} = \underline{\underline{0,5976}}$$



In der Stichprobe liegt ein  
mäßiger, positiver, linearer  
Zusammenhang zwischen  $X$  und  $Y$   
vor.