

Beschreibung des Zusammenhangs zweier metrischer Merkmale

- Streudiagramme
- Korrelationskoeffizienten
- Regression

Beispiel zu Korrelation

Beispiel 2.1:
(vgl. HEINRICH, 1998,
Veränderungen im Kieferwinkelbereich nach bilateraler sagittaler Spaltosteotomie bei Klasse II-Patienten)



- Operative Korrektur einer Klasse II-Malokklusion (mittels BSSO=bilateral sagittal split osteotomy)



Mandibulabreitenveränderung

Die Operation kann zu einer unerwünschten Veränderung der Mandibulabreite führen

Wovon wird das Ausmaß dieser Veränderung beeinflusst?

Beschreibung der Studie

Kiefer von 42 Patienten **prä- und postoperativ** anhand von Röntgenbildern vermessen.

Es wurden u.a. die präoperative Mastoidbreiten (in mm) sowie die Mandibulabreiten (in mm) prä- und postoperativ bestimmt

Die durch den Eingriff verursachte Veränderungen der Mandibulabreite ergibt sich aus der Differenz:

Veränderung =
Mandibulabreite postoperativ - Mandibulabreite präoperativ

Mastoidbreite und Mandibulabreitenveränderung

Daten der ersten 14 Patienten

| PATNR | MASTOIDBREITE PRÄOP | MANDIBULABREITENVERÄNDERUNG (postop-präop) |
|-------|---------------------|---|
| 1 | 112.31 | 4.18 |
| 2 | 108.9 | 2.20 |
| 3 | 109.89 | 3.63 |
| 4 | 114.84 | -7.81 |
| 5 | 108.35 | 2.75 |
| 6 | 106.59 | -0.55 |
| 7 | 102.85 | 3.85 |
| 8 | 99.55 | 5.81 |
| 9 | 109.45 | -2.20 |
| 10 | 113.74 | 16.39 |
| 11 | 107.58 | 1.43 |
| 12 | 112.31 | 2.31 |
| 13 | 112.75 | 5.79 |
| 14 | 103.84 | 2.20 |

Konkrete Frage:

Besteht zwischen der Mastoidbreite präoperativ und der Mandibulabreitenveränderung ein Zusammenhang?

Grafische Veranschaulichung der Daten mittels **Streudiagramm**:

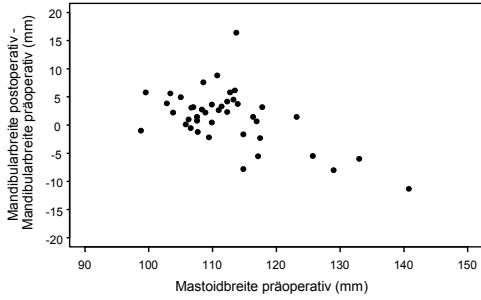
- Jeder Patient wird durch genau einen Punkt in der xy-Ebene repräsentiert.

- Die Lage dieses Punktes ist durch die beobachteten Werte in den beiden betrachteten Variablen festgelegt:

> Mastoidbreite in der x-Richtung aufgetragen,

> Mandibulabreitenveränderung in der y-Richtung aufgetragen;

Streudiagramm



Korrelationskoeffizienten

Das Ausmaß des Zusammenhanges kann mit **Korrelationskoeffizienten** quantifiziert werden.

Der Korrelationskoeffizient liegt zwischen -1 und 1 .

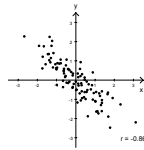
Liegt der Korrelationskoeffizient nahe bei

- 1 starke positive Korrelation.
- 0 Keine Korrelation
- 1 starke negative Korrelation.

Korrelationskoeffizienten

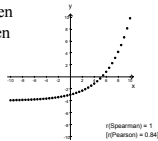
-Pearson

ist ein quantitatives Maß für die Beziehung zwischen zwei stetigen metrischen Merkmalen und misst den linearen Anteil des Zusammenhangs.

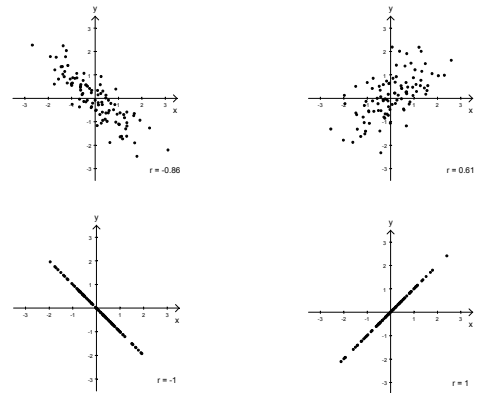


-Spearman

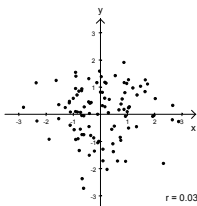
ist ein quantitatives Maß für die Beziehung zwischen zwei ordinalen Merkmalen und misst den monotonen Anteil des Zusammenhangs.



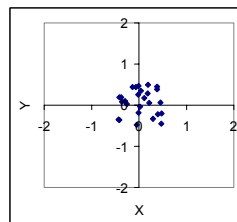
Korrelationskoeffizient nach Pearson



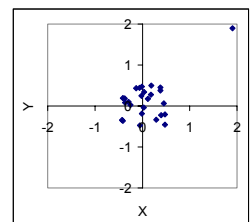
Korrelationskoeffizient nach Pearson



Korrelationskoeffizient nach Pearson Abhängigkeit von Ausreißern



$r=0.002$



$r=0.59$

Berechnung des Korrelationskoeffizienten nach Pearson

Gegeben sind die Merkmale X und Y für n Beobachtungseinheiten:

| X | Y |
|-------|-------|
| x_1 | y_1 |
| x_2 | y_2 |
| x_3 | y_3 |
| x_4 | y_4 |
| ... | ... |
| x_n | y_n |

Berechnung des Korrelationskoeffizienten nach Pearson

Der Korrelationskoeffizient nach Pearson wird mittels folgender Formel berechnet:

$$r = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}}$$

s_x^2 die Varianz des Merkmals X,

s_y^2 die Varianz des Merkmals Y,

s_{xy} die **Kovarianz** der Merkmale X und Y

s_{xy} beschreibt die gemeinsame Streuung der x- und der y-Werte.

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\left(\sum_{i=1}^n x_i y_i \right) - n \bar{x} \bar{y} \right)$$

Berechnung des Korrelationskoeffizienten nach Pearson

Fortsetzung Beispiel 2.1:

Die Stichprobenvarianz der Mastoidbreite beträgt

$$s_x^2 = 70,68 \text{ mm}^2,$$

Die Stichprobenvarianz der Mandibulabreitenveränderung

beträgt $s_y^2 = 24,48 \text{ mm}^2$,

Die Stichprobenkovarianz beträgt

$$s_{xy} = \frac{1}{41} ((11231 \cdot 4,18 + 10890 \cdot 2,20 + \dots - 11715 \cdot 5,55 + 10626 \cdot 0,99) - 42 \cdot 11229 \cdot 1,43) = -23,03$$

und für den Korrelationskoeffizienten nach Pearson ergibt sich

$$r = \frac{-23,03}{\sqrt{70,68 \cdot 24,48}} = -0,55$$

Korrelationskoeffizient r_s nach Spearman

kann verwendet werden, um eine **nicht-lineare, jedoch monotone Form** eines Zusammenhangs zwischen zwei Merkmalen zu quantifizieren.

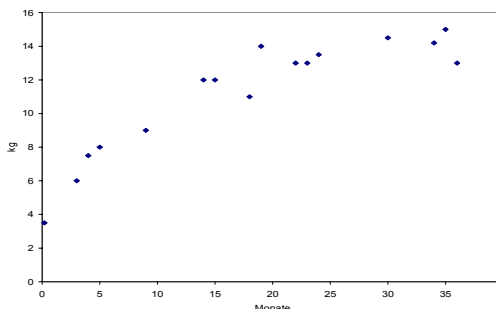
Anstatt der Originalwerte gehen die Rangzahlen in die Berechnung ein:

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n^3 - n}$$

n die Anzahl der vorliegenden Wertepaare,

D_i die Rangdifferenz des i -ten Wertepaares

Alter und Gewicht bei Kindern bis 36 Monaten



Berechnung des Korrelationskoeffizienten nach Spearman

| Kind | Monate | Gewicht |
|------|--------|---------|
| 1 | 9 | 9 |
| 2 | 5 | 8 |
| 3 | 4 | 7.5 |
| 4 | 3 | 6 |
| 5 | 22 | 13 |
| 6 | 24 | 13.5 |
| 7 | 35 | 15 |
| 8 | 18 | 11 |
| 9 | 23 | 13 |
| 10 | 36 | 13 |
| 11 | 34 | 14.2 |
| 12 | 0.2 | 3.5 |
| 13 | 30 | 14.5 |
| 14 | 15 | 12 |
| 15 | 14 | 12 |
| 16 | 19 | 14 |

Bestimmen der Rangzahlen

1. Sortieren nach dem Alter.

| Kind | Monate | Ränge (Monate) |
|------|--------|----------------|
| 12 | 0.2 | 1 |
| 4 | 3 | 2 |
| 3 | 4 | 3 |
| 2 | 5 | 4 |
| 1 | 9 | 5 |
| 15 | 14 | 6 |
| 14 | 15 | 7 |
| 8 | 18 | 8 |
| 16 | 19 | 9 |
| 5 | 22 | 10 |
| 9 | 23 | 11 |
| 6 | 24 | 12 |
| 13 | 30 | 13 |
| 11 | 34 | 14 |
| 7 | 35 | 15 |
| 10 | 36 | 16 |

Bestimmen der Rangzahlen

2. Sortieren nach dem Gewicht.

| Kind | Gewicht | Ränge | Mittränge |
|------|---------|-------|-----------|
| 12 | 3.5 | 1 | 1 |
| 4 | 6 | 2 | 2 |
| 3 | 7.5 | 3 | 3 |
| 2 | 8 | 4 | 4 |
| 1 | 9 | 5 | 5 |
| 8 | 11 | 6 | 6 |
| 15 | 12 | 7 | 7.5 |
| 14 | 12 | 8 | 7.5 |
| 5 | 13 | 9 | 10 |
| 9 | 13 | 10 | 10 |
| 10 | 13 | 11 | 10 |
| 6 | 13.5 | 12 | 12 |
| 16 | 14 | 13 | 13 |
| 11 | 14.2 | 14 | 14 |
| 13 | 14.5 | 15 | 15 |
| 7 | 15 | 16 | 16 |

Berechnung des Korrelationskoeffizienten nach Spearman

| Kind | Ränge (Monate) | Ränge (Gewicht) | D _i | D _i ² |
|--------|----------------|-----------------|----------------|-----------------------------|
| 1 | 5 | 5 | 0 | 0 |
| 2 | 4 | 4 | 0 | 0 |
| 3 | 3 | 3 | 0 | 0 |
| 4 | 2 | 2 | 0 | 0 |
| 5 | 10 | 10 | 0 | 0 |
| 6 | 12 | 12 | 0 | 0 |
| 7 | 15 | 16 | -1 | 1 |
| 8 | 8 | 6 | 2 | 4 |
| 9 | 11 | 10 | 1 | 1 |
| 10 | 16 | 10 | 6 | 36 |
| 11 | 14 | 14 | 0 | 0 |
| 12 | 1 | 1 | 0 | 0 |
| 13 | 13 | 15 | -2 | 4 |
| 14 | 7 | 7.5 | -0.5 | 0.25 |
| 15 | 6 | 7.5 | -1.5 | 2.25 |
| 16 | 9 | 13 | -4 | 16 |
| Summe: | | | 64.5 | |

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n^3 - n}$$

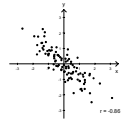
$$= 1 - \frac{6 \cdot 64,5}{16^3 - 16}$$

$$= 0,905$$

Korrelationskoeffizienten

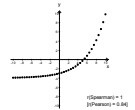
➤ Pearson

- Beschreibt lineare Zusammenhänge
- Ist für metrische Daten geeignet
- Kann durch Ausreißer beeinflusst werden.

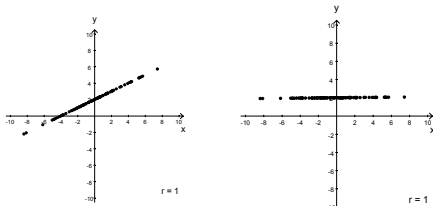


➤ Spearman

- Beschreibt monotone Zusammenhänge
- Ist ähnlich dem Korrelationskoeffizienten nach Pearson wenn ein linearer Zusammenhang besteht
- Ist für ordinalskalierte Daten geeignet
- Wird durch Ausreißer nicht beeinflusst.



Korrelationskoeffizienten



• Der Korrelationskoeffizient gibt keine direkte Auskunft über die Steigung der Geraden.

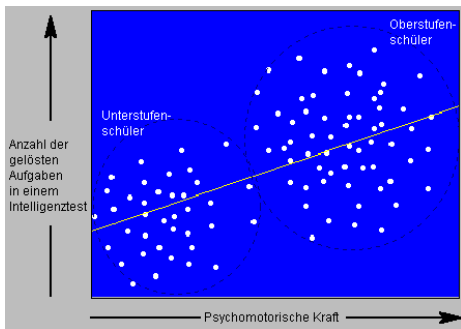
• Folglich ist ein im statistischen Sinn starker Zusammenhang nicht unbedingt ein medizinisch relevanter Zusammenhang.

Korrelationskoeffizienten

➤ Eine statistische Korrelation

- bedeutet nicht notwendigerweise einen relevanten Zusammenhang
- bedeutet nicht notwendigerweise einen kausalen Zusammenhang:
Es kann einen dritten Faktor geben, der die beiden anderen bestimmt: Bei Kindern ist etwa Gewicht und die Schnelligkeit beim Laufen positiv korreliert.
- kann auch auf einer Scheinkorrelation beruhen

Scheinkorrelation

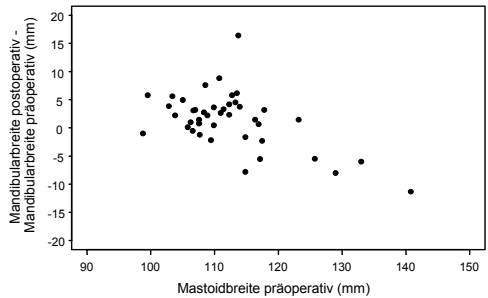


Quelle: <http://www.phil.uni-sb.de/~jakobs/seminar/vpl/expost/kausal.htm>

$r=0.6$

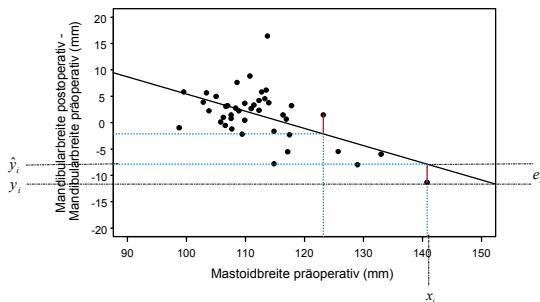
Bivariate deskriptive Verfahren

Regression



Bivariate deskriptive Verfahren

Regression



Bivariate deskriptive Verfahren

Regression

Man wählt jene Gerade als Regressionsgerade für die gilt, dass die **Quadratsumme der Residuen** am kleinsten ist.

d.h.

$$SQR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 \quad \text{ist zu minimieren!}$$

Bivariate deskriptive Verfahren

Regression

Geradengleichung allgemein: $y = k \cdot x + d$

Regressionsgerade:

$$\hat{y}_i = b_0 + b_1 \cdot x_i$$

wobei die Steigung b_1 und der Achsenabschnitt b_0 folgendermaßen berechnet werden können:

$$b_1 = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y}{s_x} \quad b_0 = \bar{y} - b_1 \cdot \bar{x}$$

Bivariate deskriptive Verfahren

Regression

Lineare Einfachregression für Beispiel 2.1:

gesucht: $\hat{y}_i = b_0 + b_1 \cdot x_i$

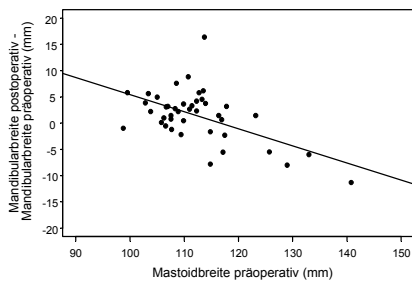
$$\text{also } b_1 = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y}{s_x} \quad \text{und } b_0 = \bar{y} - b_1 \cdot \bar{x} ;$$

wissen: $s_x^2 = 70.68 \text{ MPa}^2$, $s_y^2 = 24.48 \text{ MPa}^2$ und $r = -0,55$;

$$b_1 = -0,55 \cdot \frac{\sqrt{24,48}}{\sqrt{70,68}} = -0,32 \quad b_0 = 1,43 - (-0,32) \cdot 112,29 = 37,36$$

Regressionsgerade: $\hat{y}_i = 37,36 - 0,32 \cdot x_i$

Bivariate deskriptive Verfahren Regression



Bivariate deskriptive Verfahren Regression

Folgerungen

- **Prognose:** Ein Patienten mit einer präoperativen Mastoidbreite von 100 mm hat im Mittel eine Mandibularbreitenveränderung (post-prä) von:

$$37,36 - 0,32 \cdot 100 = +5,36 \text{ mm.}$$

- **Wie y von x abhängt:** Hat Patient A eine um 1 mm größere präoperative Mastoidbreite als Patient B, dann hat im Mittel Patient A eine um 0,32 mm geringere Mandibularbreitenveränderung als Patient B.

- Der **Schwerpunkt** der Punktwolke (\bar{x} , \bar{y}) liegt auf der Regressionsgeraden (gilt allgemein!);