

Korrelation

statistik-online.ch

Eine **Korrelation** beschreibt den Zusammenhang zwischen zwei Variablen. Sie zeigt, ob hohe Werte einer Variable tendenziell mit hohen oder tiefen Werten einer anderen Variable einhergehen.

1 Grundidee

Korrelationen beschreiben Richtung und Stärke eines Zusammenhangs.

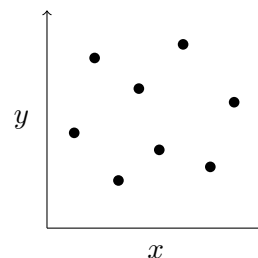
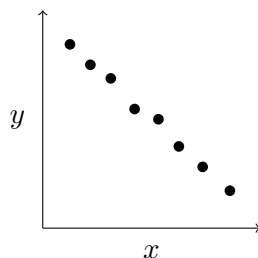
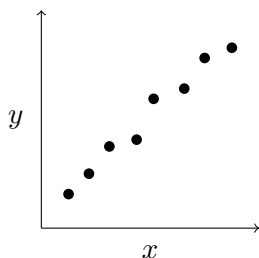
Typische Fragen sind:

- Hängen Lernzeit und Prüfungsergebnis zusammen?
- Hängen Körpergrösse und Gewicht zusammen?
- Hängen Alter und Reaktionszeit zusammen?
- Ist der Zusammenhang positiv, negativ oder praktisch nicht vorhanden?

Wichtig: Eine Korrelation beweist keine Ursache-Wirkungs-Beziehung.

2 Streudiagramm

Ein Streudiagramm ist der erste Schritt. Jeder Punkt steht für eine Person, ein Objekt oder eine Beobachtung.



positive Korrelation negative Korrelation kaum lineare Korrelation

- **Positiv:** Wenn x grösser wird, wird y tendenziell grösser.
- **Negativ:** Wenn x grösser wird, wird y tendenziell kleiner.
- **Nahe 0:** Es gibt kaum einen linearen Zusammenhang.

3 Pearson-Korrelation

Die Pearson-Korrelation misst den **linearen Zusammenhang** zwischen zwei metrischen Variablen.

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

Symbol	Bedeutung
r	Korrelationskoeffizient nach Pearson
x_i, y_i	einzelne Werte der beiden Variablen
\bar{x}, \bar{y}	Mittelwerte der beiden Variablen
Σ	Summe über alle Beobachtungen

Der Wert von r liegt immer zwischen -1 und $+1$.

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

4 Interpretation von r

Wert von r	Interpretation
$r \approx +1$	starker positiver linearer Zusammenhang
$r \approx -1$	starker negativer linearer Zusammenhang
$r \approx 0$	kaum linearer Zusammenhang
$r > 0$	je grösser x , desto tendenziell grösser y
$r < 0$	je grösser x , desto tendenziell kleiner y

Grobe Faustregel:

$ r $	Stärke
0.00 bis 0.10	kaum Zusammenhang
0.10 bis 0.30	schwacher Zusammenhang
0.30 bis 0.50	mittlerer Zusammenhang
ab 0.50	starker Zusammenhang

Diese Grenzen sind nur Orientierung. In manchen Fachgebieten gelten bereits kleinere Korrelationen als wichtig.

5 Bestimmtheitsmass r^2

Das Bestimmtheitsmass r^2 beschreibt, welcher Anteil der Streuung in einer Variable durch einen linearen Zusammenhang mit der anderen Variable erklärt werden kann.

$$r^2 = r \cdot r$$

Beispiel:

$$r = 0.80 \quad r^2 = 0.80^2 = 0.64$$

Das bedeutet: Etwa 64% der Streuung können durch den linearen Zusammenhang beschrieben werden. Die restlichen 36% werden nicht durch diesen einfachen linearen Zusammenhang erklärt.

6 Pearson oder Spearman?

Korrelation	Verwendung	Typische Daten
Pearson	linearer Zusammenhang	metrische Variablen
Spearman	monotoner Zusammenhang	ordinale Daten oder Rangdaten

Spearman-Korrelation wird verwendet, wenn die Werte als Ränge betrachtet werden oder wenn der Zusammenhang nicht gut linear, aber trotzdem geordnet ist.

7 Voraussetzungen und Grenzen

- Pearson setzt einen ungefähr linearen Zusammenhang voraus.
- Ausreisser können r stark beeinflussen.
- $r = 0$ bedeutet nur: kein linearer Zusammenhang. Ein nicht-linearer Zusammenhang kann trotzdem vorhanden sein.
- Korrelation bedeutet nicht Kausalität.
- Eine dritte Variable kann beide Variablen beeinflussen.

8 Beispiel

Eine Lehrperson untersucht den Zusammenhang zwischen Lernzeit und Prüfungspunkten.

Person	x : Lernzeit	y : Punkte	$x_i - \bar{x}$	$y_i - \bar{y}$
1	1	54	-2.5	-11
2	2	62	-1.5	-3
3	3	59	-0.5	-6
4	4	68	0.5	3
5	5	72	1.5	7
6	6	75	2.5	10

$$\bar{x} = 3.5 \quad \bar{y} = 65$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = 72$$

$$\sum (x_i - \bar{x})^2 = 17.5 \quad \sum (y_i - \bar{y})^2 = 324$$

$$r = \frac{72}{\sqrt{17.5 \cdot 324}} = \frac{72}{75.30} = 0.96$$

Interpretation: Es gibt einen starken positiven linearen Zusammenhang. Personen mit mehr Lernzeit haben in diesen Daten tendenziell mehr Punkte.

$$r^2 = 0.96^2 = 0.92$$

Etwa 92% der Streuung der Punkte können durch den linearen Zusammenhang mit der Lernzeit beschrieben werden.

9 Zusammenfassung

Begriff	Bedeutung	Wichtig
Korrelation	Zusammenhang zweier Variablen	keine Kausalität
r	Pearson-Korrelation	zwischen -1 und $+1$
$r > 0$	positiver Zusammenhang	beide Variablen steigen gemeinsam
$r < 0$	negativer Zusammenhang	eine steigt, die andere sinkt
r^2	Bestimmtheitsmass	erklärter Streuungsanteil
Spearman	Rangkorrelation	ordinale oder monotone Daten

Aufgaben: Korrelation

1 – Interpretation

Interpretieren Sie die folgenden Korrelationswerte:

Beispiel	Korrelationswert	Interpretation
a) Lernzeit und Punkte	$r = 0.72$	
b) Alter und Reaktionsgeschwindigkeit	$r = -0.58$	
c) Schuhgrösse und Lieblingsfarbe	$r = 0.03$	
d) Stress und Schlafdauer	$r = -0.31$	

2 – Pearson-Korrelation berechnen

Berechnen Sie die Pearson-Korrelation für die folgenden Daten.

Person	1	2	3	4
x	1	2	3	4
y	2	3	5	6

- Berechnen Sie \bar{x} und \bar{y} .
- Berechnen Sie r .
- Interpretieren Sie das Ergebnis.

3 – Bestimmtheitsmass

Eine Studie findet eine Korrelation von $r = -0.60$ zwischen Stress und Schlafdauer.

- Berechnen Sie r^2 .
- Interpretieren Sie r und r^2 .

4 – Multiple Choice

Welche Aussagen sind korrekt? (Mehrfachantworten möglich)

- Eine Korrelation von $r = 0.90$ beweist, dass x die Ursache von y ist.
- Bei $r = -0.75$ besteht ein starker negativer linearer Zusammenhang.
- Pearson eignet sich besonders für lineare Zusammenhänge.
- $r = 0$ beweist, dass es gar keinen Zusammenhang zwischen zwei Variablen gibt.
- Ausreisser können die Pearson-Korrelation stark beeinflussen.

Lösungen: Korrelation

Lösung zu Aufgabe 1

Beispiel	Korrelationswert	Interpretation
a)	$r = 0.72$	starker positiver Zusammenhang
b)	$r = -0.58$	starker negativer Zusammenhang
c)	$r = 0.03$	praktisch kein linearer Zusammenhang
d)	$r = -0.31$	mittlerer negativer Zusammenhang

Lösung zu Aufgabe 2

$$\bar{x} = \frac{1 + 2 + 3 + 4}{4} = 2.5 \quad \bar{y} = \frac{2 + 3 + 5 + 6}{4} = 4$$

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	2	-1.5	-2	3.0	2.25
2	3	-0.5	-1	0.5	0.25
3	5	0.5	1	0.5	0.25
4	6	1.5	2	3.0	2.25
				7.0	5.0

$$\sum (y_i - \bar{y})^2 = 4 + 1 + 1 + 4 = 10$$

$$r = \frac{7}{\sqrt{5 \cdot 10}} = \frac{7}{7.07} = 0.99$$

Interpretation: Es gibt einen sehr starken positiven linearen Zusammenhang.

Lösung zu Aufgabe 3

$$r^2 = (-0.60)^2 = 0.36$$

Interpretation von r : Es gibt einen mittleren bis starken negativen Zusammenhang. Wenn Stress höher ist, ist die Schlafdauer tendenziell tiefer.

Interpretation von r^2 : Etwa 36% der Streuung können durch den linearen Zusammenhang beschrieben werden.

Lösung zu Aufgabe 4

- ✗ Falsch – Korrelation beweist keine Kausalität.
- ✓ Richtig – $r = -0.75$ ist stark und negativ.
- ✓ Richtig – Pearson misst lineare Zusammenhänge.
- ✗ Falsch – $r = 0$ bedeutet nur keinen linearen Zusammenhang.
- ✓ Richtig – Ausreisser können r stark verändern.