



KORRELATION UND SCHEINKORRELATION

Sarah Pestkowski
Institut für Mathematik



Was ist Korrelation?

Eine **Korrelation** beschreibt eine Beziehung zwischen zwei oder mehreren Merkmalen oder Funktionen. Um die Stärke der Beziehung zwischen den Merkmalen zu erfassen, haben Statistiker ein besonderes Werkzeug: den **Korrelationskoeffizient**. [1]

Der Korrelationskoeffizient kann zwischen -1 und +1 liegen und gibt an, ob es einen linearen Zusammenhang zwischen zwei Datenreihen gibt. Bei +1 lassen sich beide Datenreihen exakt in Deckung bringen, bei -1 besteht ein umgekehrter Zusammenhang. Liegt der Korrelationskoeffizient bei Null, so gibt es gar keinen linearen Zusammenhang (es könnte aber durchaus nichtlineare Zusammenhänge geben!). [2] [3]

Korrelationskoeffizient [2]

Die zwei Merkmale x_i und y_i einer Stichprobe sollen in einen Zusammenhang gebracht werden. Wir wollen also den Korrelationskoeffizienten für die zwei gegebene Datenreihen berechnen. Diese Rechnung ist recht einfach durchführbar.

Dazu wird zunächst die sogenannte Stichprobenkovarianz berechnet:

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

Dabei bezeichnen \bar{x} und \bar{y} die Mittelwerte der Stichproben. Mit s_x und s_y sind die Standardabweichungen der Datenreihen gemeint.

Anschließend kann der Korrelationskoeffizient nach folgender Formel berechnet werden:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

Veranschaulichung des Korrelationskoeffizienten

In den nachfolgenden Abbildungen werden verschiedene Werte des Korrelationskoeffizienten veranschaulicht. Das erste Bild zeigt einen starken linearen Zusammenhang, während das zweite Bild einen starken umgekehrten linearen Zusammenhang zeigt. Im dritten Bild liegen alle Werte verstreut, es ist keine Korrelation vorhanden. Das vierte Bild macht besonders deutlich, dass es durchaus einen Zusammenhang zwischen den Merkmalen x und y geben kann, der aber vom Korrelationskoeffizienten nicht erkannt wird, da er nur lineare Zusammenhänge zeigt. [2]

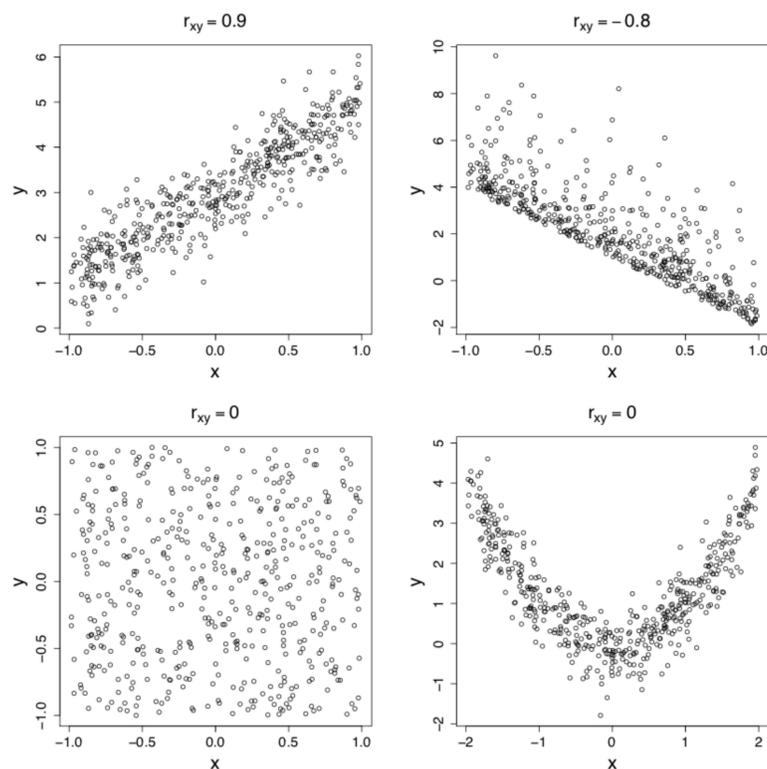


Fig. 1: Veranschaulichung verschiedener Korrelationskoeffizienten [2]

Gibt es da wirklich einen Zusammenhang?

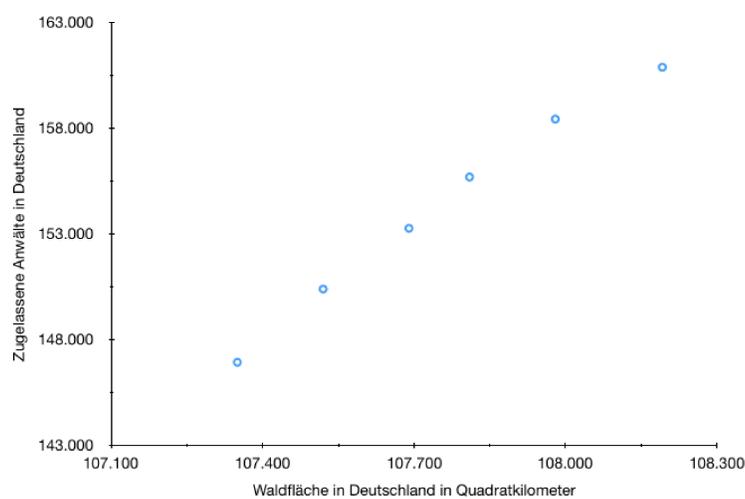


Fig. 2: Waldfläche in Deutschland in Quadratkilometern und zugelassene Anwälte in Deutschland jeweils in den Jahren 2008 bis 2013
Korrelationskoeffizient: $r = 0,9956$

[3] [4]

Scheinkorrelationen

Aber trotz scheinbarer graphischer Abhängigkeit zwischen zwei Merkmalen - wie in dem Bild links zu erkennen ist - kann nicht immer darauf geschlossen werden, dass ein kausaler Zusammenhang vorliegt. Denn es kann weitere Größen geben, die bei dem Korrelationskoeffizienten unberücksichtigt bleiben. [3]

Eine **Scheinkorrelation** beschreibt einen Zusammenhang zwischen zwei Merkmalen, zwischen denen kein kausaler Zusammenhang besteht, sondern lediglich eine zufällige oder indirekte Beziehung.

Ohne einen **kausalen Zusammenhang** erfolgt die Zuordnung zwischen zwei Merkmalen x und y willkürlich und ohne fundierte Begründung. [1]

Teste deine Intuition!

Liegen in den separaten Grafiken Korrelationen oder Scheinkorrelationen vor?

Literatur

- [1] https://de.wikipedia.org/wiki/Korrelation_bzw._/Scheinkorrelation
- [2] Edgar Steiger. ITB HU Berlin. *Skript zur Vorlesung Biostatistik*.
- [3] <https://scheinkorrelation.jimdo.com>
- [4] <https://www.brak.de/statistiken/>