

Kapitel II

Korrelationsanalyse

D. 2. 1. (Gegenstand)

Die Korrelationsanalyse untersucht, ob ein Zusammenhang zwischen zwei metrischen Variablen *besteht*, wie *stark* dieser Zusammenhang ist und welche "*Richtung*" er hat.

B. 2. 1.

Der Einfachheit halber gehen wir im Weiteren von einer linearen Beziehung aus.

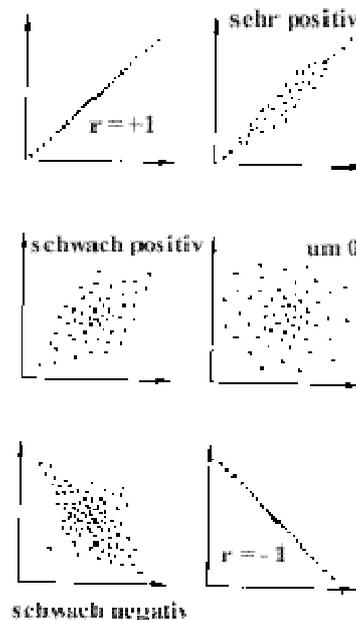
B. 2. 2. (Analysemethoden)

Es gibt zwei Methoden zur Korrelationsanalyse:

1. Mit Hilfe eines Streudiagramms
2. Formelmäßig durch den Pearson-Korrelationskoeffizienten.

B. 2. 3. (Streuungsdiagramme)

Mit Hilfe eines Streudiagramms lässt sich lediglich *grob* eine Vorstellung über die Korrelation zwischen zwei Variablen gewinnen.



D. 2. 2. (Bravais-Pearson-Korrelationskoeffizient)

Die Merkmale X und Y seien kardinalskaliert¹. Als *Bravais-Pearson-Korrelationskoeffizient* bzw. *Produkt-Moment-Korrelationskoeffizient* bezeichnet man

¹ Sollen auf der Basis des Korrelationskoeffizienten Hypothesentests durchgeführt werden, ist es zudem erforderlich, dass die Variablen in der Grundgesamtheit gemeinsam (zweidimensional) normalverteilt sind. Für Variablen, die diese Voraussetzungen nicht erfüllen, aber mindestens Ordinalskalenniveau besitzen, stehen die beiden Korrelationskoeffizienten *Kendalls Tau-b* und *Spearman's Rho* zur Verfügung.

$$r = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \cdot \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Dabei wird vorausgesetzt, dass weder alle x_i - Werte noch alle y_i - Werte gleich sind.

B. 2. 4. (Eigenschaften)

1.

Es gilt

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

2.

$r < 0$ bedeutet einen gegenläufigen Zusammenhang; $r > 0$ einen gleichsinnigen.

3.

$r = 0$ zeigt keinen Zusammenhang und $r = 1$ einen ausschließlichen Zusammenhang.

4.

Je näher $|r|$ an 1 liegt, desto stärker ist der Zusammenhang.

5.

r ist dimensionslos.

6.

Ein Vertauschen der Variablen ändert nicht den Messwert.

B. 2. 5

In der Formel zur Berechnung des Korrelationskoeffizienten nach Bravais-Pearson ist der Ausdruck

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

die geschätzte Kovarianz der Variablen x und y und misst die Stärke und Richtung des linearen Zusammenhangs zwischen den Variablen in Form eines nicht normierten Maßes.

Die Ausdrücke

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{und} \quad \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

sind die geschätzten Standardabweichungen s_x und s_y der Variablen x und y . Damit hat man

$$r = \frac{S_{xy}}{\sqrt{S_x \cdot S_y}}$$

B. 2. 6.

Die Formel zur Berechnung des Bravais-Pearson-Korrelationskoeffizienten lässt sich auch in folgender vereinfachten Form dargestellt werden:

$$r = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2}}$$

B. 2. 7.

Es muss davor gewarnt werden, aus der mit Hilfe eines Korrelationskoeffizienten gemessenen statistischen Korrelation zwischen zwei Variablen auf das Bestehen eines Kausalzusammenhangs zu schließen. Zwei voneinander unabhängige Variable können eine statistische Korrelation aufweisen, weil z.B. eine dritte Variable sowohl auf die eine als auch auf die andere Variable wirkt und sich dadurch beide in gleiche Richtung verändern. Bei Vorliegen einer statistisch gemessenen Korrelation ohne Vorliegen eines Kausalzusammenhangs spricht man von *Scheinkorrelation*. Ein in der Literatur gern zitiertes Beispiel für eine Scheinkorrelation ist der gemessene positive und in mittlerer Größenordnung liegende Korrelationskoeffizient für den Zusammenhang zwischen der Anzahl der gezählten Störche und der Anzahl der Geburten in einer Region.

B. 2. 8. (Statistischer Test)

Soll statistisch getestet werden, ob ein linearer Zusammenhang zwischen den Variablen x und y für die Grundgesamtheit besteht, also die Hypothese geprüft werden, ob der unbekannte Korrelationskoeffizient der Grundgesamtheit ρ sich signifikant von Null unterscheidet, so bedarf es spezieller *Annahmen*:

- i) Die gemeinsame (bivariate) Verteilung der Variablen ist normalverteilt,
- ii) Die vorliegende Stichprobendatei ist per Zufall zustande gekommen.

Unter diesen Annahmen kann der Korrelationskoeffizient nach folgenden Schritten getestet werden:

Schritte des statistischen Tests:

1. Formuliere die Null- und die Alternativhypothese

$$H_0 : \rho=0 \qquad H_1 : \rho \neq 0$$

(Die Alternativhypothese wird dabei je nach Erwartung über die Richtung des Zusammenhangs unterschiedlich formuliert. Hat man keinerlei Erwartung über die Richtung des Zusammenhangs, so gilt: $H_1 : \rho \neq 0$. Es handelt sich dann um einen zweiseitigen Test. Erwartet man, dass die Variablen sich in gleicher Richtung verändern, so wird der positive Zusammenhang mit $H_1 : \rho > 0$ formuliert. Bei

Erwartung eines negativen Zusammenhangs gilt entsprechend $H_1 : \rho < 0$. In diesen Fällen handelt es sich um einseitige Tests.)

2. Berechne die Prüfgröße:

$$t_{\text{prüf}} = r \sqrt{\frac{n-2}{1-r^2}}$$

3. Suche $t_{n-2;\alpha}$ in der Tabelle der t-Verteilung.

4. Entscheidung:

Lehne H_0 ab, falls $|t_{\text{prüf}}| > t_{n-2;\alpha}$ gilt; sonst gibt es keinen Grund, H_0 abzulehnen.

(Der letzte Fall heißt: es liegt ein statistisch gesicherter Zusammenhang zwischen den beiden Merkmalen.)

BS. 2. 1.

Die nachfolgende Tabelle zeigt die Kurse zweier Aktien X und Y an 9 aufeinander folgenden Börsentagen:

Tag	1	2	3	4	5	6	7	8	9
Aktie X	5	6	11	8	13	8	10	16	13
Aktie Y	8	7	9	10	11	10	11	12	12

Berechnen und interpretieren Sie hierzu den entsprechenden Bravais-Pearson-Koeffizienten.

Lösung:

Arbeitstabelle

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
5	8	-5	-2	10	25	4
6	7	-4	-3	12	16	9
11	9	1	-1	-1	1	1
8	10	-2	0	0	4	0
13	11	3	1	3	9	1
8	10	-2	0	0	4	0
10	11	0	1	0	0	1
16	12	6	2	12	36	4
13	12	3	2	6	9	4
90	90	0	0	42	104	24

$$\bar{x} = \frac{90}{9} = 10, \quad \bar{y} = \frac{90}{9} = 10,$$

$$r = \frac{\frac{42}{9}}{\sqrt{\frac{104}{9} \cdot \frac{24}{9}}} = 0.841273288 \approx 0.84.$$

Es handelt sich um einen relativ starken gleichsinnigen Zusammenhang.

Statistischer Test:

1. Formuliere die Null- und die Alternativhypothese

$$H_0 : \rho=0 \qquad H_1 : \rho \neq 0$$

2. Berechne die Prüfgröße:

$$t_{\text{prüf}} = r \sqrt{\frac{n-2}{1-r^2}} = 0.84 \cdot \sqrt{\frac{9-2}{1-0.84^2}} \approx 4.096$$

3. $t_{n-2;\alpha} = t_{7;0.05} = 2.365$

4. Entscheidung:

$$t_{\text{prüf}} \approx 4.096 > 2.365 = t_{7;0.05}$$

Die Nullhypothese wird abgelehnt. Es gibt also einen Zusammenhang zwischen den Kursen der Aktien X und Y .

(Letzte Aktualisierung: 05.02.2012)