

Produkt-Moment-Korrelation (1)

- Einführung I -

- Kennziffer zur Beschreibung des **linearen** Zusammenhangs zwischen zwei Variablen X und Y.
- Beschreibt die **Richtung** und die **Enge** des Zusammenhangs im Sinne von „je..., desto...“ oder „wenn..., dann...“-Aussagen.
Bsp. Je geringer die Konfliktkompetenz, desto ausgeprägter ist die Gewalteinstellung.
- Symbol: r oder $r_{x,y}$; wird auch **Pearson-Korrelationskoeffizient** genannt
- Setzt intervallskalierte Daten voraus
(um Rückschlüsse auf die gemessenen Merkmale ziehen zu können)

Produkt-Moment-Korrelation (2)

- Einführung II -

Fragestellung in der Korrelationsanalyse:

1. Besteht ein Zusammenhang zwischen zwei Merkmalen?
[Gehen hohe Werte in X eher mit hohen (oder niedrigen) Werten in Y einher?]
[Wie spiegeln sich Veränderungen in X in den Ausprägungen von Y wider?]
2. Wie ausgeprägt ist dieser Zusammenhang?

Produkt-Moment-Korrelation (3)

- Exkurs: Unterscheidung deterministischer und stochastischer Zusammenhang -

⇒ **Deterministischer oder funktionaler Zusammenhang**

Y läßt sich **genau** aus X vorhersagen.

Unterschiede in Y korrespondieren perfekt mit Unterschieden in X.

Bei linearen Zusammenhängen gilt: $y_i = a + b_{yx} \cdot x_i$

Bsp.: Telefonrechnung = Grundgebühr + 0.10 Cent · Einheiten

⇒ **Stochastischer oder probabilistischer Zusammenhang**

Y läßt sich zwar aus X vorhersagen, jedoch ist die Ausprägung von Y noch von anderen Variablen außer X abhängig.

Unterschiede in Y korrespondieren zwar mit Unterschieden in X, aber es treten im Einzelfall **Ungenauigkeiten bei der Vorhersage** auf.

Bei linearen Zusammenhängen gilt: $y_{ij} = a + b_{yx} \cdot x_i + e_{ij}$

Bsp.: Gewalteinstellung = b · Elterngewalt + **Fehler**
(Fehler sind z. B. Normorientierung, Empathie etc.)

Produkt-Moment-Korrelation (4)

- Herleitung der Produkt-Moment-Korrelation über die Kovarianz -

Die Kovarianz ist **maßstabsabhängig**, d.h. sie ist **nicht** invariant gegenüber linearen Transformationen.

$$s_{xy} = \text{COV}_{(x,y)} = \frac{\sum_i (X_i - \bar{x}) \cdot (Y_i - \bar{y})}{N}$$

Ein **maßstabsunabhängiges** Maß $r_{x,y}$ für den Zusammenhang zweier intervallskalierter Variablen X und Y erhält man, wenn man die Kovarianz durch das Produkt der beiden Streuungen dividiert. Die Korrelation ist also die standardisierte Kovarianz:

$$r_{x,y} = \frac{\text{COV}_{(x,y)}}{s_x \cdot s_y}$$

Produkt-Moment-Korrelation (5)

- Definition der Produkt-Moment-Korrelation I -

Herleitung der Maßstabsunabhängigkeit der Produkt-Moment-Korrelation.

Bei linearer Transformation der Variablen X und Y in $x' = a_1 + b_1 x$ und $y' = a_2 + b_2 y$ gilt:

$$\begin{aligned} s_{x'} &= |b_1| \cdot s_x \\ s_{y'} &= |b_2| \cdot s_y \\ \text{COV}_{(x',y')} &= b_1 \cdot b_2 \cdot \text{COV}_{(x,y)} \end{aligned}$$

Eingesetzt in die Gleichung der Produkt-Moment-Korrelation ergibt sich:

$$r_{x',y'} = \frac{\text{COV}_{(x',y')}}{s_{x'} \cdot s_{y'}} = \frac{b_1 \cdot b_2 \cdot \text{COV}_{(x,y)}}{|b_1| \cdot s_x \cdot |b_2| \cdot s_y} = \frac{\text{COV}_{(x,y)}}{s_x \cdot s_y} = r_{x,y}$$

Produkt-Moment-Korrelation (6)

- Definition der Produkt-Moment-Korrelation II -

Ersetzt man $\text{cov}_{(x,y)}$ durch die Gleichung für die Kovarianz, ergibt sich die Korrelation als arithmetisches Mittel des Produktes korrespondierender z-Werte.

$$r_{x,y} = aM(z_x \cdot z_y)$$

Herleitung →

Da die Mittelwerte von z_x und $z_y = 0$ sind, können z-Werte als Abweichungswerte vom mittleren z-Wert interpretiert werden. Damit entspricht die Korrelation der Kovarianz z-standardisierter Werte.

$$r_{x,y} = \text{COV}_{(z_x, z_y)}$$

$$\begin{aligned} r_{x,y} &= \frac{\text{COV}_{(x,y)}}{s_x \cdot s_y} \\ &= \frac{\sum_{i=1}^N (X_i - \bar{x}) \cdot (Y_i - \bar{y})}{N \cdot s_x \cdot s_y} \\ &= \frac{1}{N} \cdot \sum_{i=1}^N \frac{(X_i - \bar{x}) \cdot (Y_i - \bar{y})}{s_x \cdot s_y} \\ &= \frac{1}{N} \cdot \sum_{i=1}^N z_{x_i} \cdot z_{y_i} = aM(z_{x_i} \cdot z_{y_i}) \end{aligned}$$

Produkt-Moment-Korrelation (7)

- Die Skala der Produkt-Moment-Korrelation -

- ⇒ Der Wertebereich der Produkt-Moment-Korrelation (r) liegt zwischen -1 und +1.
 - $r = +1$: Der Zusammenhang zwischen X und Y ist perfekt **positiv** linear:
Je größer X, desto größer Y.
Zwischen X und Y besteht ein **gleichläufiger** Zusammenhang.
 - $r \approx 0$: Es besteht kein linearer Zusammenhang.
 - $r = -1$: Der Zusammenhang zwischen Y und X ist perfekt **negativ** linear:
Je kleiner X, desto größer ist Y.
Zwischen X und Y besteht ein **gegenläufiger** Zusammenhang.
- ⇒ Grobe Richtlinie
 - ab einer Stichprobengröße von $n=20$ kann eine Korrelation von
 - $|r| = 0.3$ als bedeutsam
 - $|r| = 0.5$ als gut
 - $|r| = 0.7$ als hoch bezeichnet werden.
- ⇒ Allgemein ist die Interpretation von Korrelationen aber immer kontextabhängig.

Produkt-Moment-Korrelation (8)

- Anschauliche Darstellungen von Korrelationen -

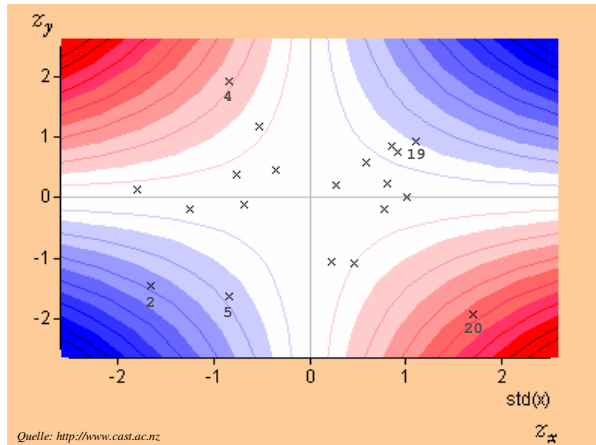
Die Produkt-Moment Korrelation r ist eine Kennziffer (Korrelationskoeffizient), die sowohl die Enge als auch die Richtung des **Zusammenhangs z-standardisierter Messwerte widerspiegelt**. Dabei beeinflussen Wertepaare den Korrelationskoeffizienten um so stärker, je weiter die Produkte der entsprechenden z-Werte von Null verschieden sind.

Im Folgenden werden zwei Streudiagramme (*scatterplots*) von Paaren z-standardisierter Werte mit dazugehörigen Tabellen dargestellt (pro Objekt / Person existiert je ein Messwert der X- und der Y-Variable).

Die roten und blauen Flächen der Diagramme entsprechen unterschiedlichen Wertebereichen der Produkte $z_x \cdot z_y$. Je dunkler das **Blau** ist, in dem ein Messwertpaar (Kreuzchen) liegt, desto größer ist sein **positiver Beitrag** zu r , und je dunkler das **Rot**, desto größer sein **negativer Beitrag**.

Produkt-Moment-Korrelation (9)

- Grafische Darstellung einer Korrelation von $r = .001$ -



Wenn die Summe der positiven und negativen Produkte fast gleich ist, ist Korrelation fast Null.

Produkt-Moment-Korrelation (10)

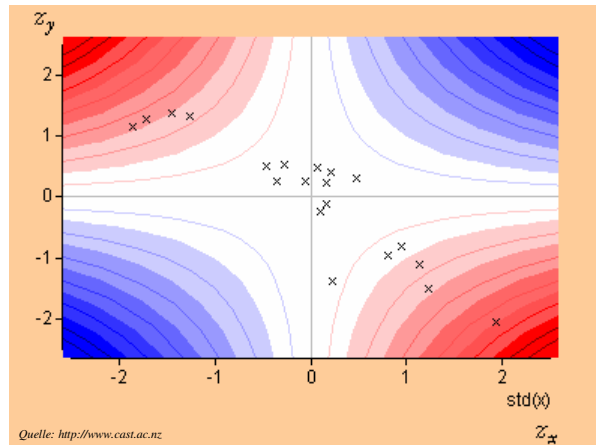
- Tabellarische Darstellung einer Korrelation von $r = .001$ -

Wertepaar	Z_x	Z_y	$Z_x \cdot Z_y$
1	-1.80	0.13	-0.2340
2	-1.66	-1.45	2.4070
3	-1.36	-0.18	0.2250
4	-0.84	1.92	-1.6128
5	-0.84	-1.64	1.3776
6	-0.76	0.39	-0.2964
7	-0.68	-0.10	0.0680
8	-0.53	1.18	-0.6254
9	-0.36	0.45	-0.1620
10	0.22	-1.05	-0.2310
11	0.28	0.21	0.0588
12	0.47	-1.09	-0.5123
13	0.59	0.59	0.3481
14	0.77	-0.19	-0.1463
15	0.80	0.24	0.1920
16	0.85	0.85	0.7225
17	0.91	0.76	0.6916
18	1.01	0.01	0.0101
19	1.10	0.94	1.0340
20	1.70	-1.94	-3.2980

$$r = \frac{\sum_{i=1}^n Z_{xi} \cdot Z_{yi}}{n} = \frac{.0165}{20} = .001$$

Produkt-Moment-Korrelation (11)

- Grafische Darstellung einer Korrelation von $r = -.855$ -



Liegen die meisten Messwertpaare in der roten Fläche, ist die Korrelation negativ.

Produkt-Moment-Korrelation (12)

- Tabellarische Darstellung einer Korrelation von $r = -.855$ -

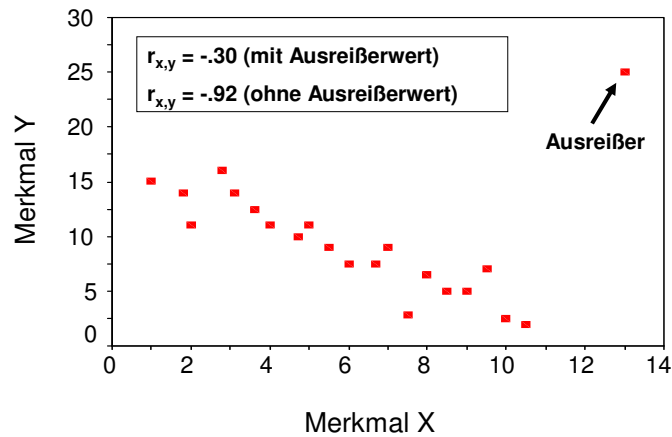
Wertepaar	z_x	z_y	$z_x \cdot z_y$
1	-1.86	1.15	-2.1390
2	-1.72	1.27	-2.1844
3	-1.45	1.38	-2.0010
4	-1.26	1.33	-1.6758
5	-0.46	0.51	-0.2346
6	-0.35	0.25	-0.0875
7	-0.28	0.54	-0.1512
8	-0.05	0.25	-0.0125
9	0.07	0.49	0.0343
10	0.10	-0.23	-0.0230
11	0.16	-0.11	-0.0176
12	0.17	0.24	0.0408
13	0.21	0.40	0.0840
14	0.23	-1.37	-0.3151
15	0.48	0.32	0.1536
16	0.80	-0.95	-0.7600
17	0.95	-0.82	-0.7790
18	1.13	-1.11	-1.2543
19	1.23	-1.50	-1.8450
20	1.93	-2.04	-3.9372

$$r = \frac{\sum_{i=1}^n z_{x,i} \cdot z_{y,i}}{n} = \frac{-17.1045}{20} = -.855$$

Produkt-Moment-Korrelation (13)

- Ausreißerwerte -

Korrelationen sind für ‚Ausreißerwerte‘ sehr anfällig:



Produkt-Moment-Korrelation (14)

- Rechenbeispiel -

x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
85	43	-15	-7	225	49	105
116	55	+16	+5	256	25	80
119	54	+19	+4	361	16	76
101	48	+1	-2	1	4	-2
72	41	-28	-9	784	81	252
93	53	-7	+3	49	9	-21
107	49	+7	-1	49	1	-7
98	51	-2	+1	4	1	-2
89	47	-11	-3	121	9	33
120	59	+20	+9	400	81	180
$\Sigma = 1000$	$\Sigma = 500$	$\bar{x} = 100$	$\bar{y} = 50$	$\Sigma = 2250$	$\Sigma = 276$	$\Sigma = 694$

$$s_x^2 = \frac{2250}{10} = 225 \Rightarrow s_x = 15 \quad \text{cov}_{(x,y)} = \frac{694}{10} = 69.4$$

$$s_y^2 = \frac{276}{10} = 27.6 \Rightarrow s_y = 5.25 \quad r_{xy} = \frac{69.4}{15 \cdot 5.25} = 0.88$$

Produkt-Moment-Korrelation (15)

- Der Rückschluss von der Korrelation der Messwerte auf die Korrelation der gemessenen Merkmale-

- Der Korrelationskoeffizient $r_{x,y}$ ist gegenüber linearen Transformationen **invariant**: $r_{x,y} = r_{x',y'}$. Werden X und Y linear transformiert, so ändert sich ihre Produkt-Moment-Korrelation $r_{x,y}$ nicht.
- Ausnahme: bei einer linearen Umpolung ändert sich die Richtung des Zusammenhangs, d.h. das Vorzeichen der P-M-K. Werden beide Variablen linear umgepolt, so heben sich die Vorzeichenwechsel gegenseitig auf.
- Bei intervallskalierten Messwerten ist die P-M-K der Messwerte gleich der P-M-K der gemessenen Merkmale.
- Liegen ordinalskalierte Daten vor, kann die P-M-K berechnet werden, der Rückschluss auf die Korrelation der gemessenen Merkmale ist dann aber **nicht** zulässig.
- Welche der beiden Korrelationen (die der Messwerte oder der Merkmale) die wichtigere ist, ist von der jeweiligen Fragestellung abhängig.