

Arbeitsvorlage zur Ermittlung eines linearen Regressionszusammenhangs zweier Merkmale Methode der kleinsten Quadrate

Unter- nehmen i	1	2	3	4	5	6	7	8	9	10	Σ
Mitarb.- anzahl x_i	3	2	4	5	8	7	5	3	2	1	
Umsatz- höhe y_i	30	20	60	70	100	80	50	40	30	20	
$x_i \cdot y_i$											
$(x_i)^2$											
$(y_i)^2$											

Ansatz: $y = b_0 + b_1 x$

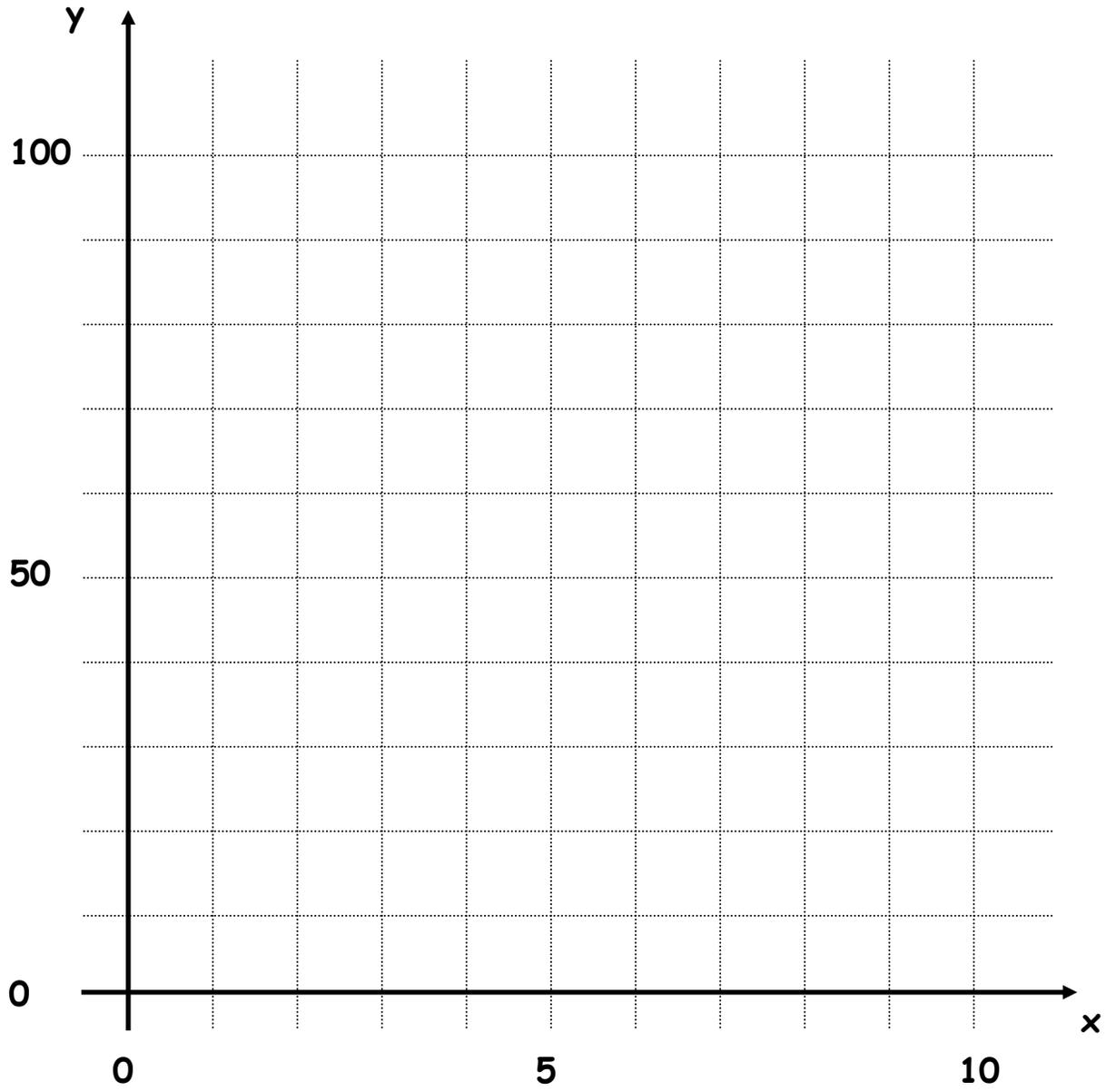
Notwendige Formeln:

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \mu_y = \frac{1}{n} \sum_{i=1}^n y_i$$

$$b_1 = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sum_{i=1}^n (x_i y_i) - n \cdot \mu_x \cdot \mu_y}{\sum_{i=1}^n (x_i^2) - n \cdot \mu_x^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i y_i) - \mu_x \cdot \mu_y}{\frac{1}{n} \sum_{i=1}^n (x_i^2) - \mu_x^2}$$

$$b_0 = \mu_y - b_1 \cdot \mu_x$$

Koordinatensystem zum Eintragen der Punkte und der Regressionsgerade



Bravais-Pearson-Korrelationskoeffizient

Der Bravais-Pearson-Korrelationskoeffizient wird zur Messung eines linearen Zusammenhangs, der sogenannten [Korrelation](#) zwischen zwei Beobachtungsgrößen verwendet. Das Datenmaterial liegt hierfür in Form von einer gepaarten Meßreihe, d.h. eines [zweidimensionalen Datensatzes](#) vor. Der Bravais-Pearson-Korrelationskoeffizient r wird mit Hilfe der [empirischen Kovarianz](#) s_{Kov} der gepaarten Meßreihe und den [Standardabweichungen](#) s_1 und s_2 der einzelnen Meßreihen mittels der Formel berechnet:

$$r = \frac{\text{empirische Kovarianz der beiden Meßreihen}}{\text{Standardabweichung der ersten Meßreihe} \cdot \text{Standardabweichung der zweiten Meßreihe}} = \frac{s_{Kov}}{s_1 s_2}$$

Da der Bravais-Pearson-Korrelationskoeffizient ein Maß für die Korrelation beider Merkmale ist, werden folgende Bezeichnungen verwendet. Die beiden Merkmale heißen

positiv korreliert, falls $r > 0$,

unkorreliert, falls $r = 0$,

negativ korreliert, falls $r < 0$.

2. Der Korrelationskoeffizient nach Bravais-Pearson

Handelt es sich bei den Variablen x und y um quantitative (kardinal meßbare) Merkmale, so kann der Grad des Zusammenhangs durch den Bravais-Pearson-Maßkorrelationskoeffizient gemessen werden.

Dieser Korrelationskoeffizient geht von den numerischen Abweichungen der beobachteten Werte von ihrem Mittelwert aus und betrachtet damit nicht nur die Differenzen zwischen Rangplätzen. Dadurch werden die in den Daten enthaltenen Informationen besser ausgewertet.

Der Korrelationskoeffizient nach Bravais-Pearson r ist definiert als:

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

Der Korrelationskoeffizient r nimmt ebenso wie ρ Werte zwischen $+1$ und -1 an, die wie ρ zu interpretieren sind.

Beispiel:

In unserem Beispiel ist der Zusammenhang zwischen Werbekosten und Absatz nach dem [Bravais-Pearson-Korrelationskoeffizient](#) zu bestimmen.

Die zur Berechnung notwendigen Daten können der Arbeitstabelle 71 entnommen werden.

$$r = \frac{647,5}{\sqrt{74\,800 \cdot 5,97}} = \frac{647,5}{668,25} = \underline{\underline{0,969}}$$

Maßzahlen für den Zusammenhang zweier Variablen, Korrelation

Die Stärke des Zusammenhanges zwischen zwei metrisch skalierten Merkmalen X und Y wird durch die gemeinsame [Variation](#) der beiden Merkmale bestimmt.

Unter [Variation](#) im Sinne der Streuungsbetrachtung wird die Abweichung der Merkmalsausprägungen von ihrem arithmetischen [Mittelwert](#) betrachtet.

Die Merkmalswerte werden in einem ersten Schritt zentriert:

$$x_i^* = x_i - \bar{x}$$

$$y_i^* = y_i - \bar{y}$$

Die gemeinsame [Variation](#) beider Merkmale ergibt sich als Produkt der Abweichungen der Beobachtungen vom [Mittelwert](#) (vgl. auch Berechnung der Kovarianz):

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

Das Ausmaß der gemeinsamen [Variation](#) wird in starkem Maße von der Maßeinheit der Merkmale und der Anzahl der Beobachtungen beeinflusst

So beträgt beispielsweise der [Mittelwert](#) des einen Merkmals 8 und der Beobachtungswert 10, der [Mittelwert](#) des anderen Merkmals 1008 und der Beobachtungswert 1260. Obwohl die Abweichung des ersten Wertes 2 und des zweiten Wertes dagegen 252 beträgt, ist die relative Abweichung vom [Mittelwert](#) bei beiden gleich 25%.

Um vergleichbare Abweichungen der Merkmale zu erreichen, wird eine Standardisierung vorgenommen:

$$(x_i - \bar{x}) / s_x$$

$$(y_i - \bar{y}) / s_y$$

Die obige Gleichung verändert sich dadurch zu:

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

Diese Produktsumme wird abschließend durch die Anzahl der Beobachtungen dividiert, um deren Einfluss zu eliminieren.

Damit ergibt sich der Bravais-Pearson-Korrelationskoeffizient, der es erlaubt, die Stärke des **linearen** Zusammenhangs zwischen zwei metrisch skalierten Merkmalen X und Y zu messen:

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

[\[bearbeiten\]](#)

Bravais-Pearson-Korrelationskoeffizient

Wie die Vereinfachung der obigen Gleichung zeigt, entspricht der Bravais-Pearson-Korrelationskoeffizient der gemeinsamen Streuung der beiden Merkmale X und Y (= Kovarianz) normiert auf das Produkt der Einzelstreuung (= [Standardabweichung](#)) der Merkmale.

Der Bravais-Pearson-Korrelationskoeffizient lässt sich auch in der folgenden Form darstellen

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

$$r_{yx} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}$$

Eigenschaften des Korrelationskoeffizienten:

- Der Korrelationskoeffizient nimmt nur Werte zwischen -1 und +1 an: $-1 \leq r_{xy} \leq +1$
- Das Vorzeichen des Korrelationskoeffizienten gibt Auskunft über die Richtung des Zusammenhangs
 - "+" entspricht einer positiven Korrelation (Proportionalität in der Streuung)
 - "-" entspricht einer negativen Korrelation (umgekehrte Proportionalität in der Streuung)
- Liegen alle Beobachtungswerte auf einer Geraden, nimmt der Korrelationskoeffizient den Wert $|1|$ an.

Je mehr sich der Korrelationskoeffizient dem Wert $|1|$ nähert, desto ausgeprägter ist ein linearer Zusammenhang zwischen den Merkmalen X und Y (analog umgekehrt).

- Sind die Merkmale X und Y voneinander unabhängig, nimmt der Korrelationskoeffizient den Wert 0 an.

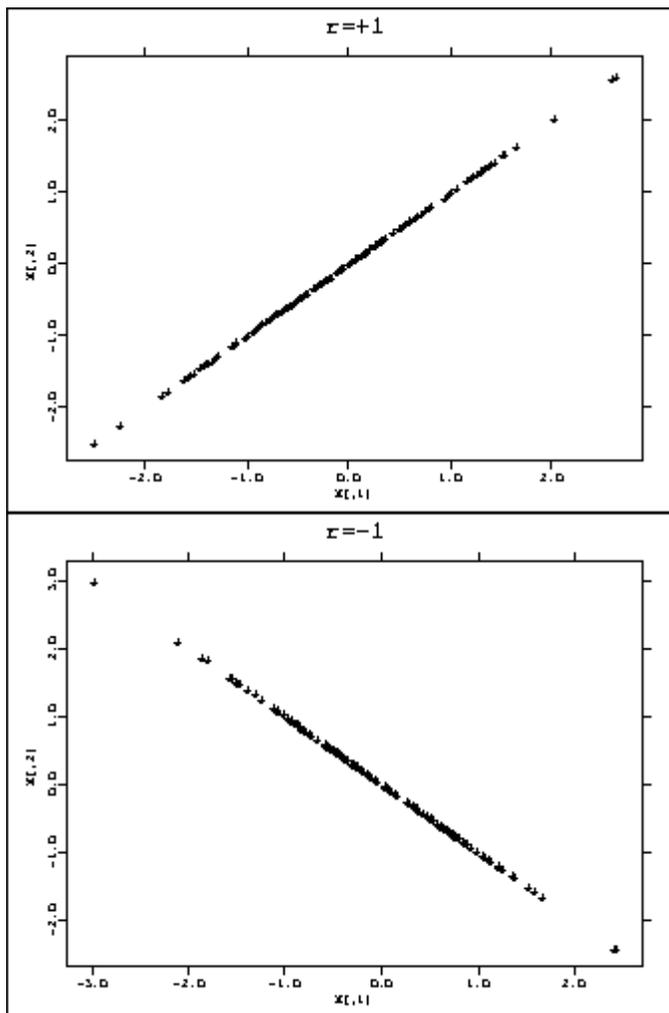
Ein Korrelationskoeffizient von 0 bedeutet demgegenüber nur, daß kein linearer Zusammenhang zwischen den Merkmalen X und Y existiert. Es ist aber durchaus möglich, daß zwischen beiden Merkmalen ein ausgeprägter nichtlinearer Zusammenhang besteht.

- Die Richtung der Beeinflussung hat keinen Einfluß auf den Wert des Korrelationskoeffizienten:

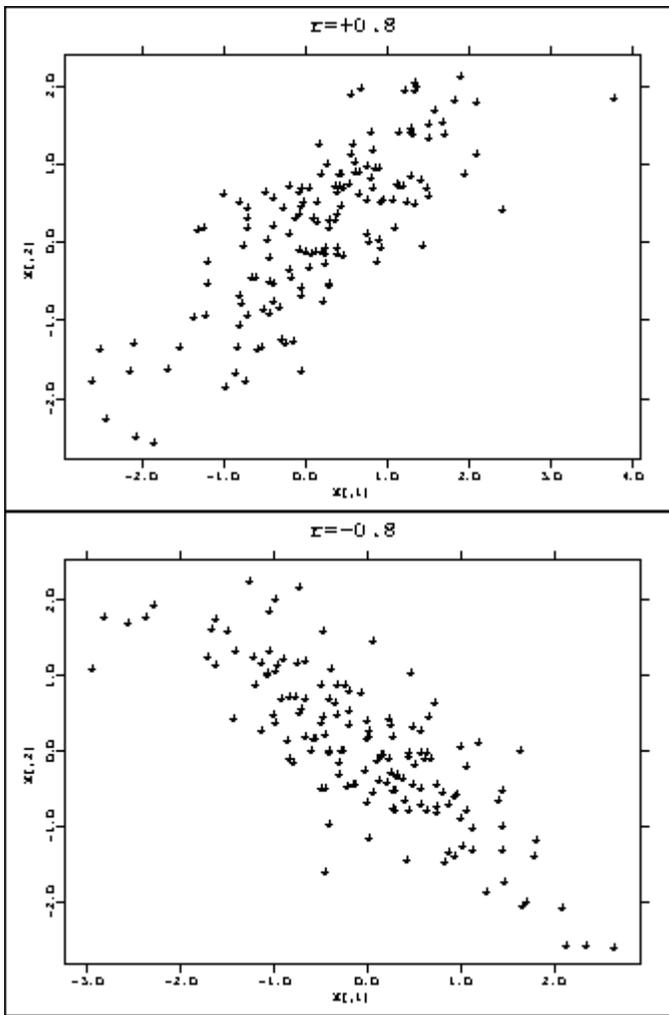
$$r_{xy} = r_{yx}$$

Zusammenhang von Korrelation und der Punktwolke der Beobachtungswerte von X und Y

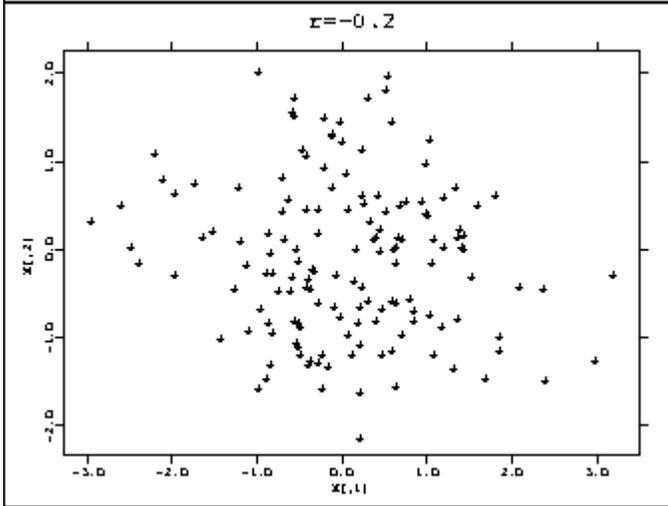
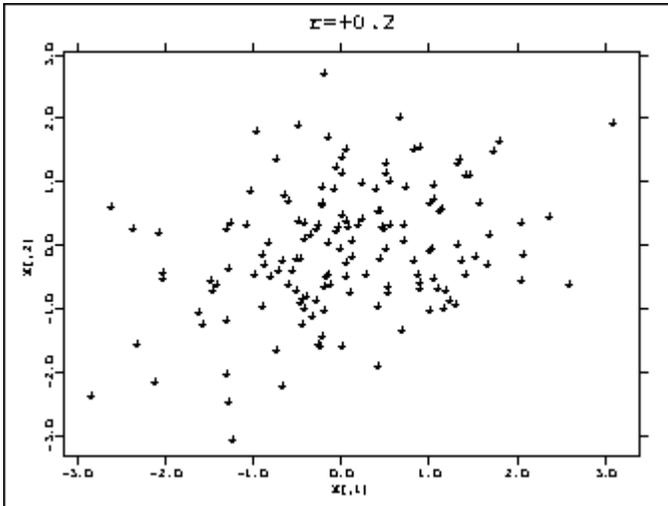
Perfekte Korrelation (Korrelationskoeffizient = $|1|$)



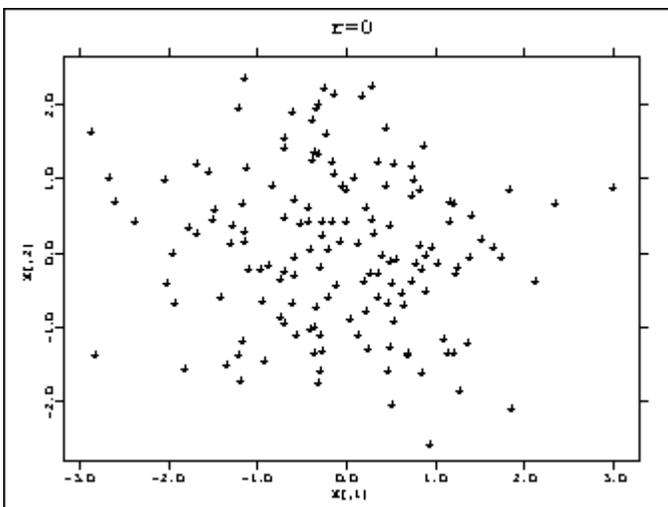
Starke Korrelation (Korrelationskoeffizient $> |0,5|$)



Schwache Korrelation (Korrelationskoeffizient $< |0,5|$)



Keine Korrelation (Korrelationskoeffizient = 0)



Eine Korrelation Null entspricht "im allgemeinen" einer kreisähnlichen Form der Punktwolke.

$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{S_{xy}}{s_x s_y},$$

wobei $s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ die empirische Kovarianz ist.

Eigenschaften der (empirischen) Kovarianz:

Satz 5.1:

- $s_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$
- $z_i = y_i + a \Rightarrow s_{xz} = s_{xy}$ die empirische Kovarianz ändert sich nicht, wenn eine Konstante zu einem Merkmal addiert wird.
- $z_i = a y_i \Rightarrow s_{xz} = a s_{xy}$
allgemein: $s_{(ax+b)(cy+d)} = a c s_{xy}$
- $|s_{xy}| \leq s_x s_y$
- $|s_{xy}| = s_x s_y \Leftrightarrow y_i = a x_i + b$ mit $a \neq 0$

Eigenschaften des Bravais-Pearson Korrelationskoeffizienten:

- $-1 \leq r_{xy} \leq 1$
- $r_{xy} = +1 \Leftrightarrow y_i = a x_i + b$ mit $a > 0$ (größte positive Korrelation)
- $r_{xy} = -1 \Leftrightarrow y_i = a x_i + b$ mit $a < 0$ (größte negative Korrelation)