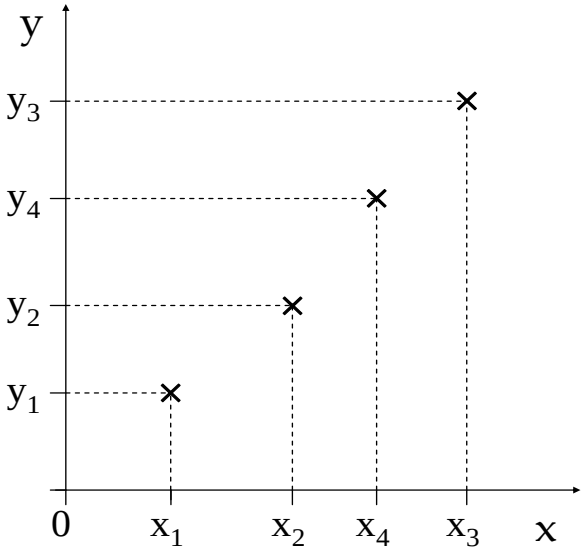
8. Korrelationsanalyse

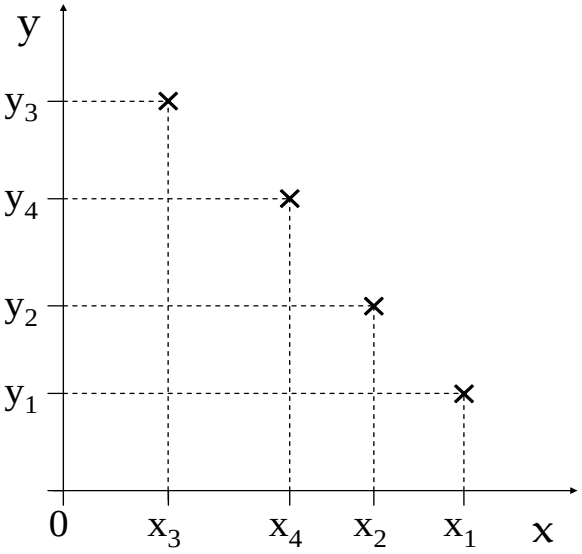
8.1 Streuungsdiagramm

- Visualisierung der gemeinsamen Variation zweier Merkmale X und Y in einem x,y-Koordinatensystem (Streuungsdiagramm)
- Anwendung bei quantitativen Merkmalen in unklassierter Form:

i	1	2	3	…	n
x_i	x_1	x_2	x_3	…	x_n
y_i	y_1	y_2	y_3	…	y_n



positiver Zusammenhang

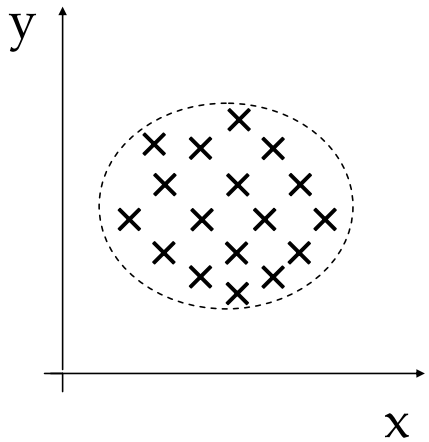


negativer Zusammenhang

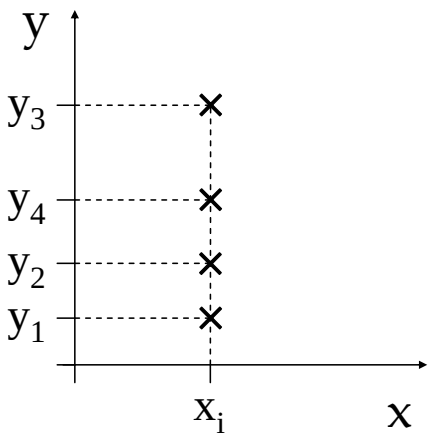
(Fehlende) Zusammenhänge im Streuungsdiagramm

- Positiver Zusammenhang: mit x ↑ folgt tendenziell y ↑
- Negativer Zusammenhang: mit x ↑ folgt tendenziell y ↓
- Kein Zusammenhang: mit x ↑ (=, ↓) folgt tendenziell y ↑, = oder ↓

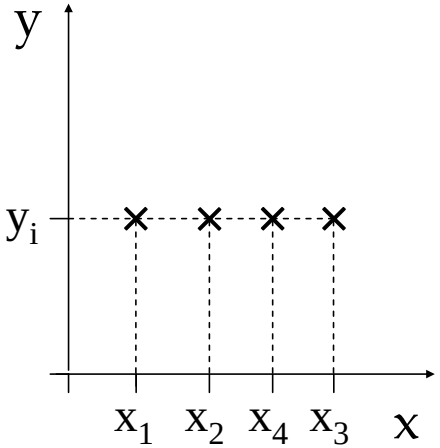
Streuungsdiagramme bei fehlendem Zusammenhang



Kreisförmige Punktewolke



Parallele zur y-Achse



Parallele zur x-Achse

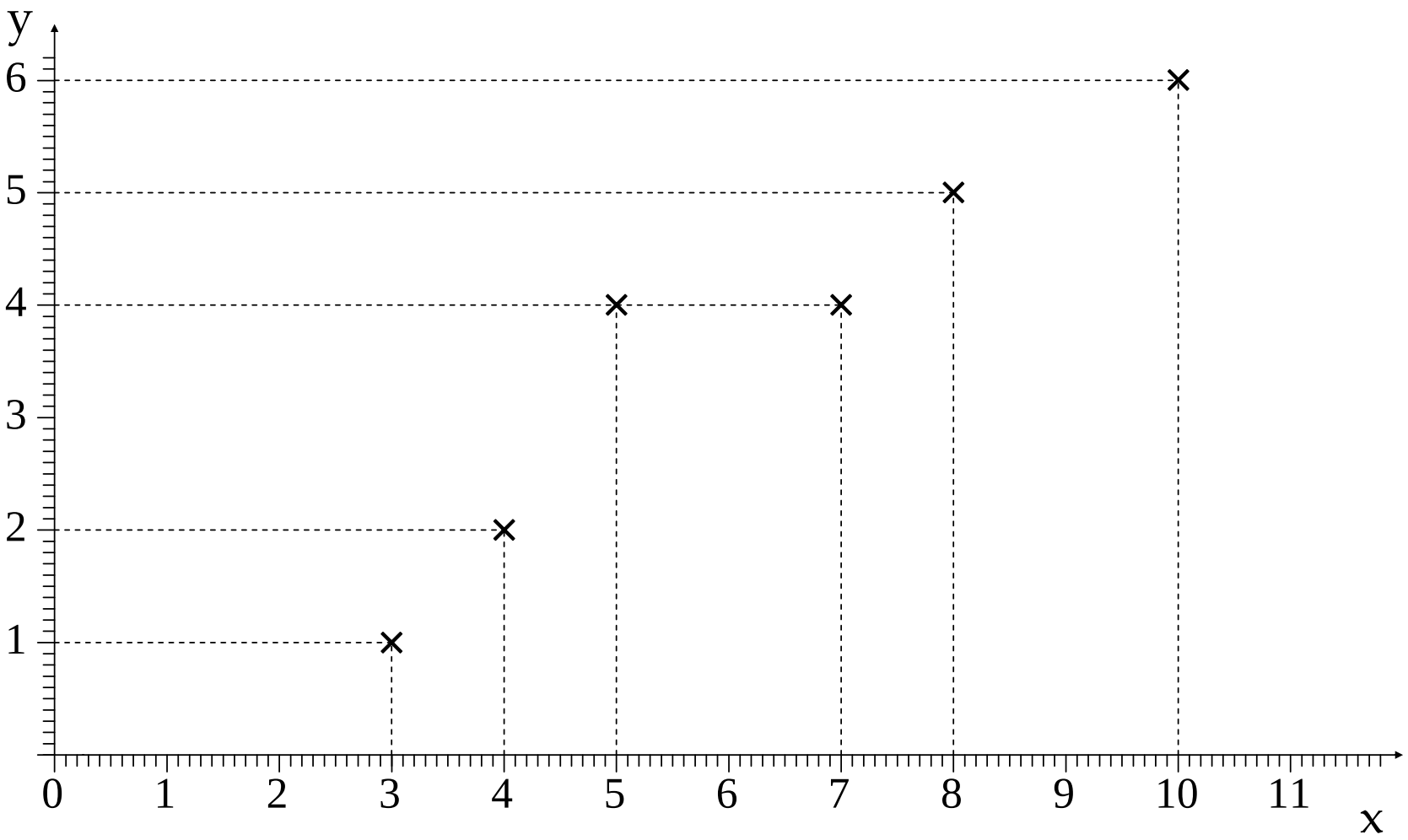
Beispiel: Werbeausgaben und Umsatz (I)

Ein Unternehmen möchte wissen, ob sich die Werbeausgaben lohnen. In diesem Fall müsste ein positiver Zusammenhang zwischen Werbeausgaben und einer Zielgröße z.B. dem Umsatz bestehen. Das Unternehmen hat die verbundenen Einzelwerte der Werbeausgaben und des Umsatzes von sechs Jahren vorliegen:

i	x_i (Werbeausgaben in 1000 €)	y_i (Umsatz in Mio. €)
1	3	1
2	5	4
3	4	2
4	7	4
5	8	5
6	10	6

Mit diesen Informationen soll ein Streuungsdiagramm gezeichnet werden. Welches Merkmal auf der x-Achse abgetragen wird, ist beliebig. Wir verwenden die Werbeausgaben als Merkmal X.

Beispiel: Werbeausgaben und Umsatz (II)



Positiver Zusammenhang: Mit zunehmenden Werbeausgaben X steigen tendenziell auch die Umsätze Y.

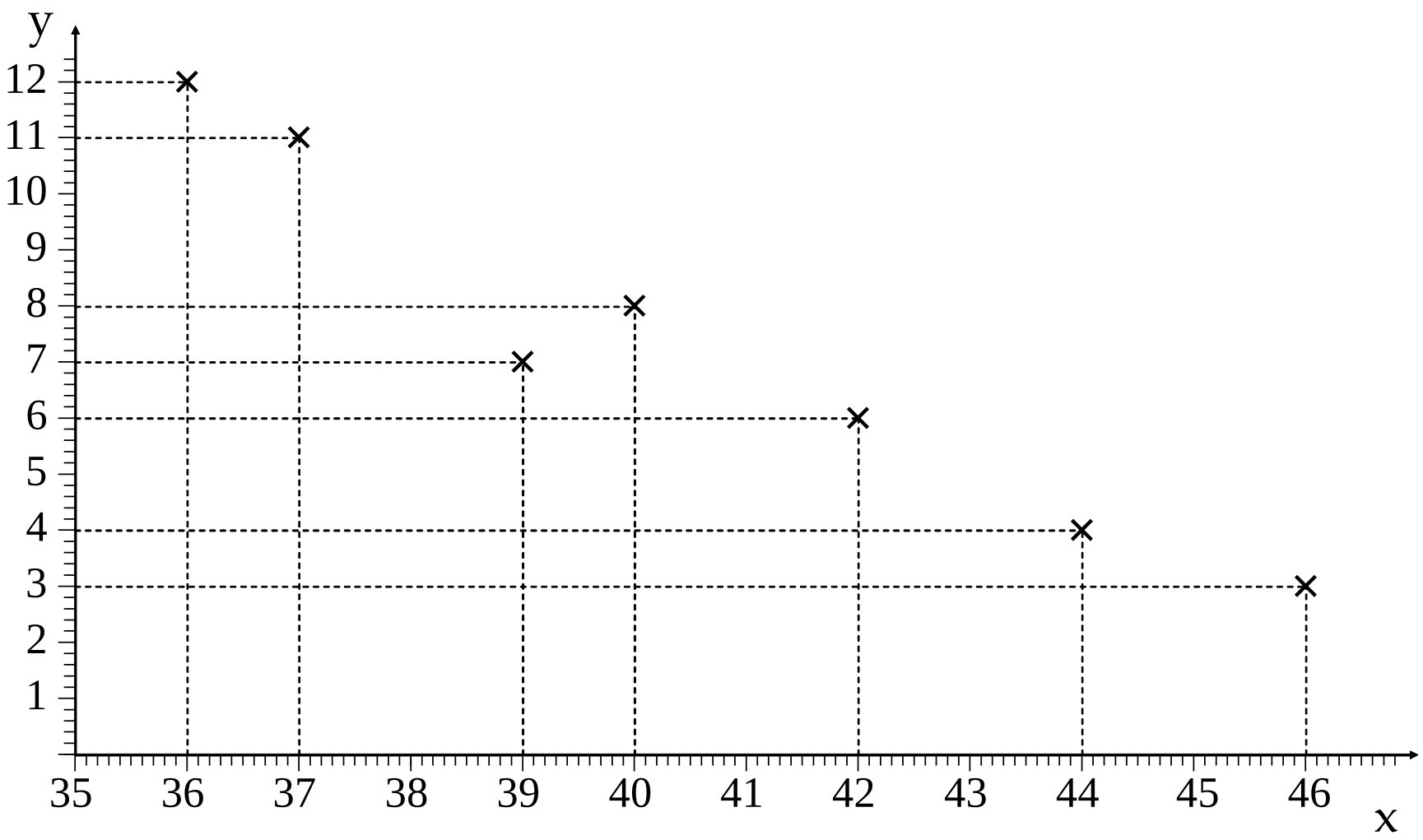
Beispiel: Heimwerkerartikel (I)

Einen Hersteller von Heimwerkerartikeln interessiert der Zusammenhang zwischen dem Verkaufspreis eines Produktes und der verkauften Stückzahl (Absatz). In sieben Testmärkten wird sein Produkt deshalb zu unterschiedlichen Preisen angeboten und der Absatz registriert.

i	x_i (Preis in €)	y_i (Absatz in Tsd. Stück)
1	44	4
2	40	8
3	42	6
4	46	3
5	36	12
6	37	11
7	39	7

Der Preis wird als Merkmal X und der Absatz als Merkmal Y verwendet. Da mit steigenden Preisen tendenziell geringere Absätze einhergehen, besteht zwischen beiden Merkmalen ein negativer Zusammenhang.

Beispiel: Heimwerkerartikel (II)

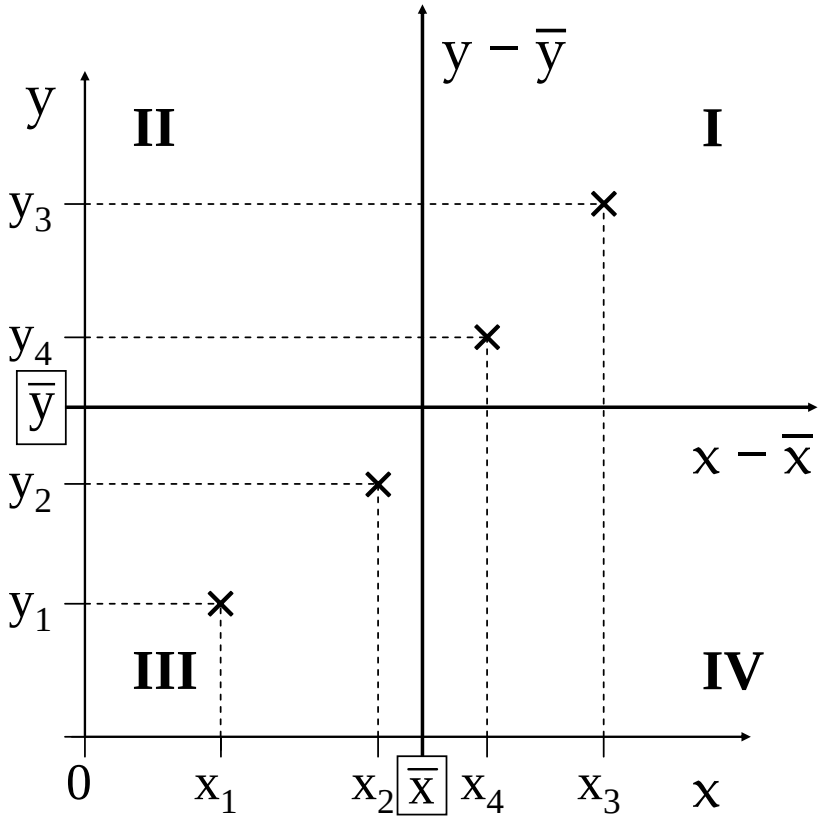


Negativer Zusammenhang: Mit steigenden Preisen gehen tendenziell geringere Umsätze einher.

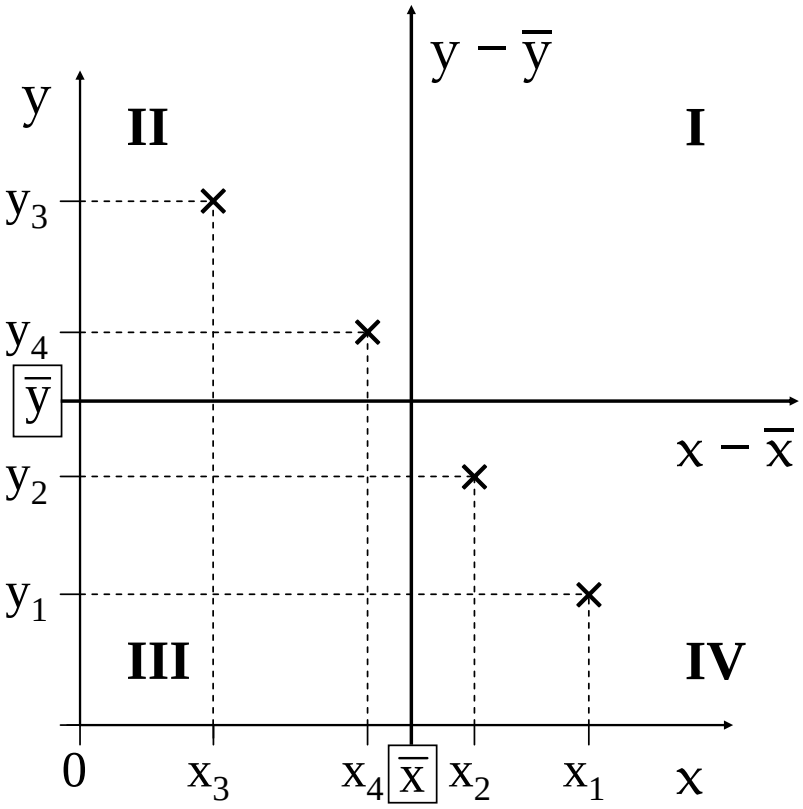
8.2 Kovarianz

- Die **Kovarianz** ist eine Maßzahl für die gemeinsame Streuung zwischen zwei Merkmalen X und Y (= Verbundstreuung)

Streuungsdiagramm mit verschobenem Koordinatensystem



Positiver Zusammenhang



Negativer Zusammenhang

Beispiel (Fortsetzung): Werbeausgaben und Umsatz (I)

- Um das neue Koordinatensystem einzuzeichnen, müssen der x- und y-Wert des neuen Ursprungs, also die beiden arithmetischen Mittel, berechnet werden.

- Im Beispiel der Werbeausgaben und der Umsätze erhalten wir:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{6} \cdot (3 + 5 + 4 + 7 + 8 + 10) = 6{,}167$$

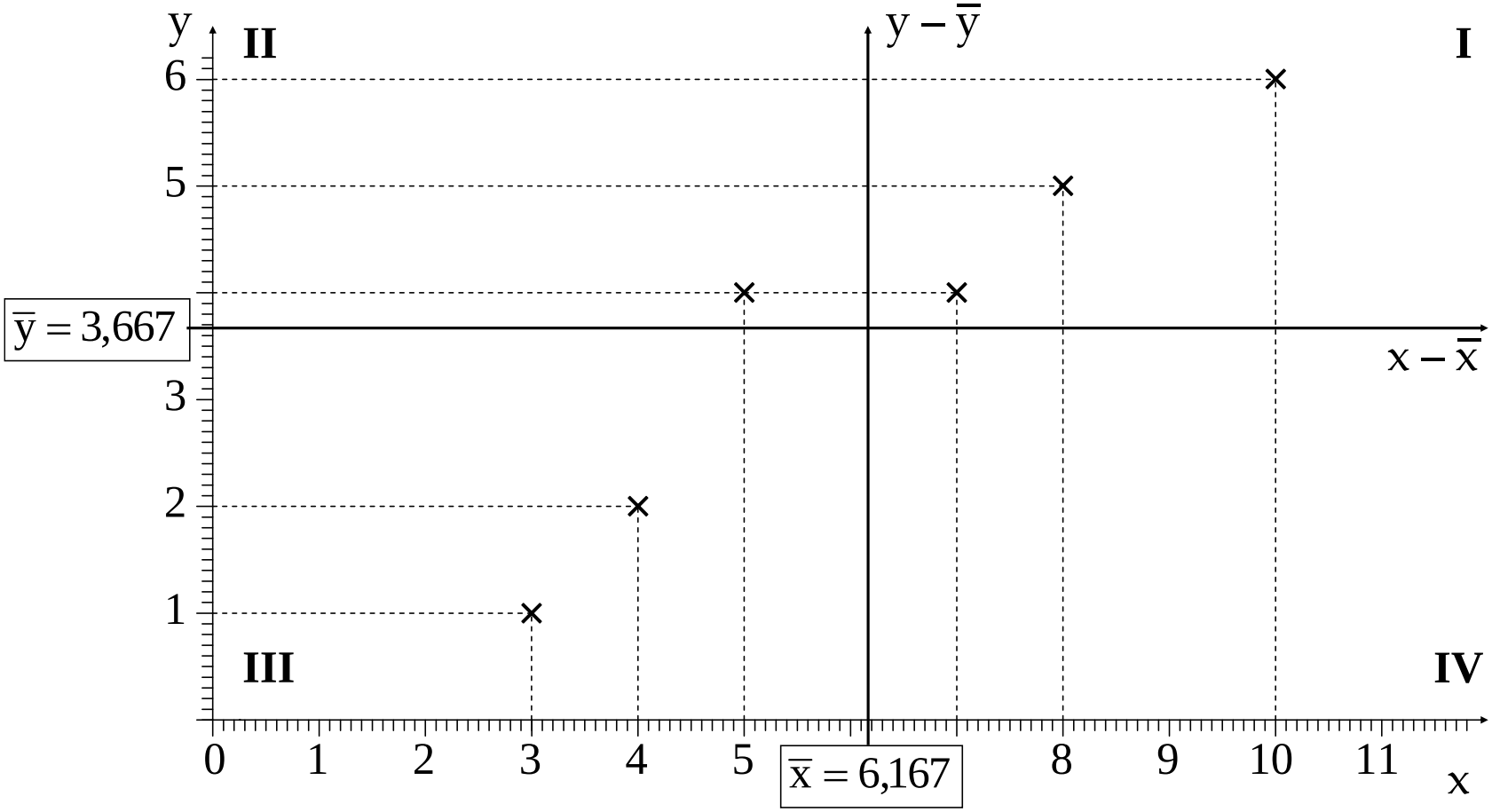
$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = \frac{1}{6} \cdot (1 + 4 + 2 + 4 + 5 + 6) = 3{,}667$$

- Nachdem das neue Koordinatensystem in das Streuungsdiagramm eingezeichnet wurde, kann festgestellt werden, welche Beobachtungen in die einzelnen Quadranten fallen:

  - **Positiver Zusammenhang:**
    Häufung der verbundenen Beobachtungen im I. und III. Quadranten

  - **Negativer Zusammenhang:**
    Häufung der Beobachtungen im II. und IV. Quadranten

Beispiel (Fortsetzung): Werbeausgaben und Umsatz (II)



Positiver Zusammenhang: Fünf Beobachtungen in Quadrant I und III, nur eine Beobachtung in Quadrant II und IV.

Berechnung der Kovarianz über Verwendung der Kreuzprodukte

- Kreuzprodukte: Beiträge der verbundenen Beobachtungen zur Kovarianz

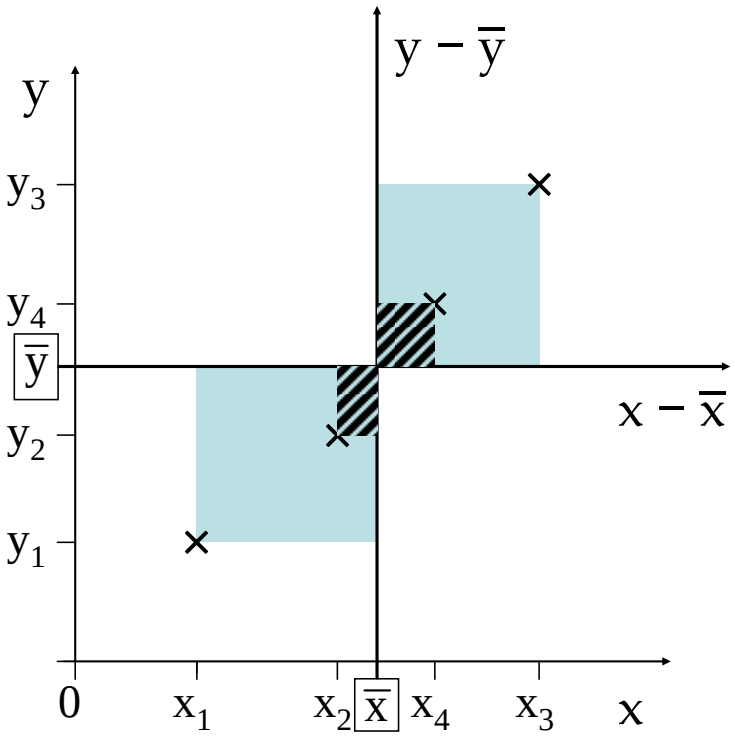
$$(x_i - \bar{x}) \cdot (y_i - \bar{y})$$

- Verbundene Beobachtungen im I. und III. Quadranten: Positives Kreuzprodukt
- Verbundene Beobachtungen im II. und IV. Quadranten: Negatives Kreuzprodukt
- Je größer die Abweichungen von den Grenzen der Quadranten, um so größer sind absolut die Kreuzprodukte (d.h. die Flächeninhalte der Rechtecke)
- Summe der Kreuzprodukte:

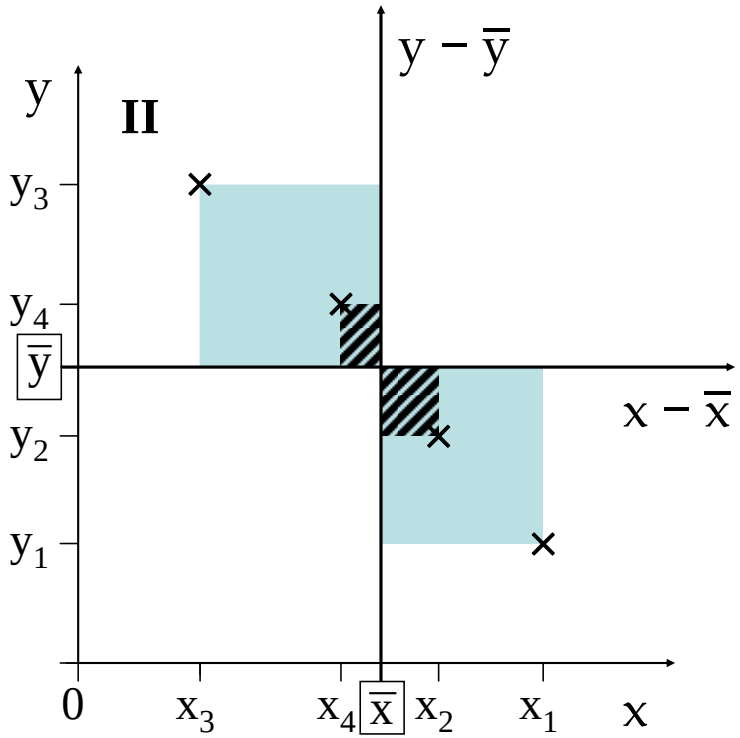
$$\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

- **Positive Summe**: Gesamtflächeninhalt der Rechtecke im I. und III. Quadranten übersteigt Gesamtflächeninhalt der Rechtecke im II. und IV. Quadranten.
- **Negative Summe**: Gesamtflächeninhalt der Rechtecke im II. und IV. Quadranten übersteigt Gesamtflächeninhalt der Rechtecke im I. und III. Quadranten.

Graphische Veranschaulichung der Kreuzprodukte



Positiver Zusammenhang



Negativer Zusammenhang

Kovarianz: Durchschnittliches Kreuzprodukt

- Die Kovarianz ergibt sich als arithmetisches Mittel der Kreuzprodukte:

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

- Interpretation der Kovarianz:
  - $s_{xy} > 0$: Positiver Zusammenhang (positive Kreuzprodukte überwiegen)
  - $s_{xy} = 0$: Kein (linearer) Zusammenhang (positive und negative Kreuz-produkte gleichen sich aus)
  - $s_{xy} < 0$: Negativer Zusammenhang (negative Kreuzprodukte überwiegen)

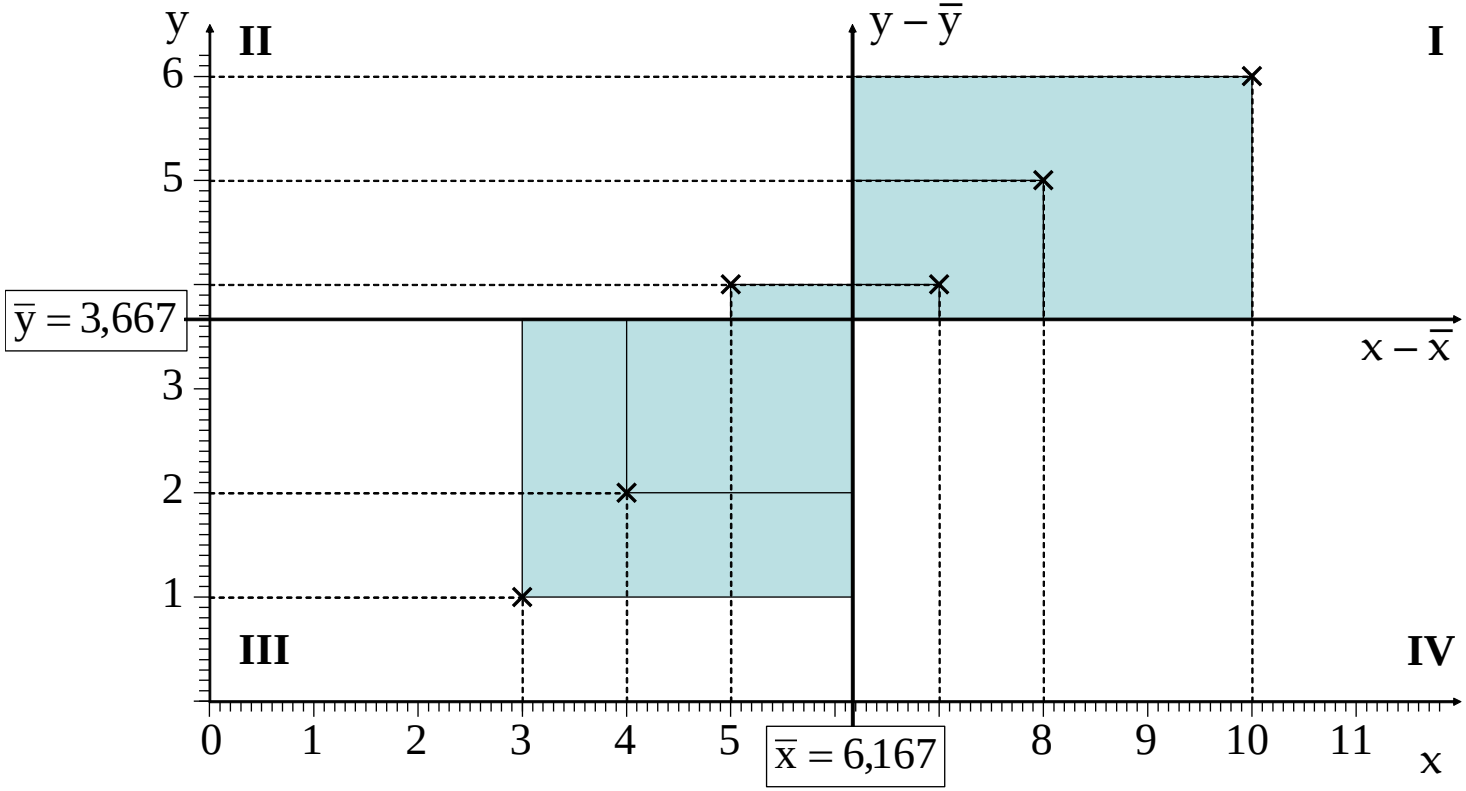
Beispiel (Fortsetzung): Werbeausgaben und Umsatz (I)

Wie hoch ist die Kovarianz für die Werbeausgaben und den Umsatz? Die arithmetischen Mittel beider Merkmale wurden bereits ermittelt ($\bar{x} = 6{,}167$, $\bar{y} = 3{,}667$). Die Summe der Kreuzprodukte wird in folgender Tabelle berechnet:

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	3	1	-3,167	-2,667	(-3,167)·(-2,667) = 8,446
2	5	4	-1,167	0,333	(-1,167)·0,333 = -0,389
3	4	2	-2,167	-1,667	(-2,167)·(-1,667) = 3,612
4	7	4	0,833	0,333	0,833·0,333 = 0,277
5	8	5	1,833	1,333	1,833·1,333 = 2,443
6	10	6	3,833	2,333	3,833·2,333 = 8,942
∑	37	22	0	0	23,331

- Negatives Kreuzprodukt: Bei verbundener Beobachtung 2
- Positive Kreuzprodukte: Bei allen anderen verbundenen Beobachtungen

Beispiel (Fortsetzung): Werbeausgaben und Umsatz (II)



- Gesamtflächeninhalt der Rechtecke im I. und III. Quadranten übersteigt Gesamtflächeninhalt im II. und IV. Quadranten $\Rightarrow$ positiver Zusammenhang zwischen Werbeausgaben und Umsatz
- Wert der Kovarianz:

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{6} \cdot 23{,}331 = 3{,}889$$

Berechnung der Kovarianz unter Verwendung des Verschiebungssatzes

- Verschiebungssatz:

$$s_{xy} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i\right) - \bar{x} \cdot \bar{y} \quad \text{oder} \quad s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}, \text{mit } \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i$$

Beispiel (Fortsetzung): Werbeausgaben und Umsatz

i	x_i	y_i	$x_i \cdot y_i$
1	3	1	3·1 = 3
2	5	4	5·4 = 20
3	4	2	4·2 = 8
4	7	4	7·4 = 28
5	8	5	8·5 = 40
6	10	6	10·6 = 60
∑	37	22	159

$$s_{xy} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i\right) - \bar{x} \cdot \bar{y} = \overline{xy} - \bar{x} \cdot \bar{y} = \frac{1}{6} \cdot 159 - 6{,}167 \cdot 3{,}667 = 3{,}886$$

Problem der Kovarianz

- Wertebereich ist nicht auf ein bestimmtes Intervall normiert
- Die Kovarianz steigt proportional mit einer Vervielfachung der x- oder y-Werte

$\Rightarrow$ Kovarianz ist noch kein Maß für die Stärke eines Zusammenhangs

Beispiel: Aktienkurse

Es wird angenommen, dass sich die gemeinsame Kursentwicklung zweier Aktien unter Verwendung ihrer Eurokurse durch eine Kovarianz s_{xy} von 20 beschreiben lässt. Bei einem Umrechnungskurs von 1,10 Dollar je 1 Euro beträgt die Kovarianz der Aktienkurse auf Dollarbasis:

$$\begin{aligned} s_{x'y'} &= \frac{1}{n}\sum_{i=1}^{n}(1{,}10 \cdot x_i - 1{,}10 \cdot \bar{x}) \cdot (1{,}10 \cdot y_i - 1{,}10 \cdot \bar{y}) \\ &= \frac{1}{n}\sum_{i=1}^{n} 1{,}10 \cdot (x_i - \bar{x}) \cdot 1{,}10 \cdot (y_i - \bar{y}) \\ &= 1{,}10^2 \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y}) = 1{,}21 \cdot s_{xy} = 1{,}21 \cdot 20 = 24{,}2 \end{aligned}$$

Man erkennt daran, dass die Kovarianz eigenständig **allein die Richtung eines Zusammenhangs** misst. Zur Messung der **Stärke** eines Zusammenhangs ist eine **Normierung erforderlich**, die diesen Transformationseffekt ausschaltet.

8.3 Korrelationskoeffizient nach Fechner

- Maß der Stärke von Zusammenhängen
- Wie bei der Kovarianz berechnet man die Kreuzprodukte $(x_i - \bar{x}) \cdot (y_i - \bar{y})$, verwendet jedoch allein ihre Vorzeichen
- Dafür Verwendung der Signum-Funktion (sgn):

$$sgn(z) = \begin{cases} -1 \text{ falls } z < 0 \\ 1 \text{ falls } z > 0 \\ 0 \text{ falls } z = 0 \end{cases}$$

- Korrelationskoeffizient nach Fechner:

$$r_F = \frac{1}{n} \cdot \sum_{i=1}^{n} sgn[(x_i - \bar{x}) \cdot (y_i - \bar{y})] \qquad \text{Normierung: } -1 \leq r_F \leq 1$$

- Alternative Berechnungsformel:

$$r_F = \frac{n^+ - n^-}{n}$$

  - n^+: Anzahl der verbundenen Beobachtungen im I. und III. Quadranten
  - n^-: Anzahl der verbundenen Beobachtungen im II. und IV. Quadranten

Art des Zusammenhangs:

- $r_F > 0$: Positiver Zusammenhang (positive Korrelation)
- $r_F \approx 0$: Geringer oder kein Zusammenhang (keine Korrelation)
- $r_F < 0$: Negativer Zusammenhang (negative Korrelation)

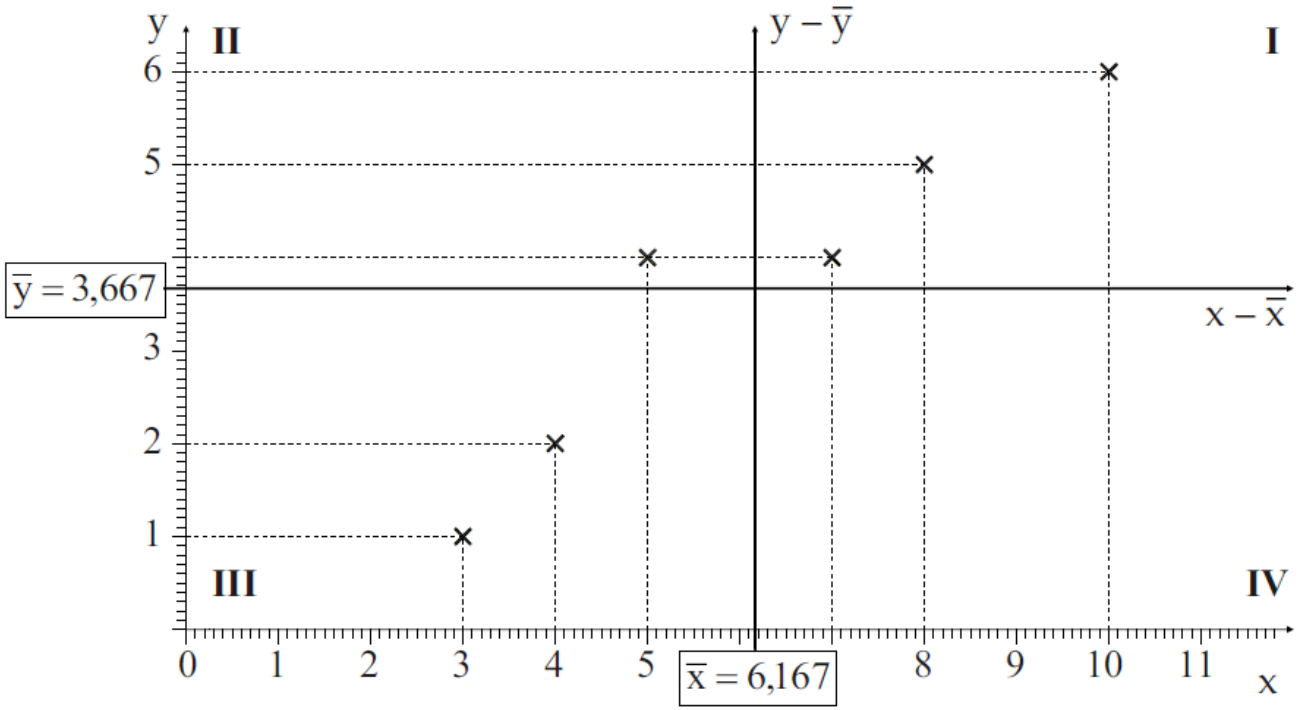
Stärke des Zusammenhangs:

- $|r_F| \geq 0{,}8$: Starke Korrelation
- $0{,}3 \leq |r_F| < 0{,}8$: Mittelstarke Korrelation
- $0 \leq |r_F| < 0{,}3$: Schwache Korrelation

Problematik des Fechnerschen Korrelationskoeffizienten:

- Die Abstände der Beobachtungen von den Grenzen der Quadranten bleiben unberücksichtigt.
- Beobachtungen mit geringen Abweichungen von den Mittelwerten werden genauso stark berücksichtigt wie Beobachtungen, die weit entfernt liegen.

Beispiel (Fortsetzung): Werbeausgaben und Umsatz



- $n^+ = 3 + 2 = 5$ (Anzahl der Beobachtungen im I. und III. Quadranten)
- $n^- = 1 + 0 = 1$ (Anzahl der Beobachtungen im II. und IV. Quadranten)
- Korrelationskoeffizient nach Fechner:

$$r_F = \frac{n^+ - n^-}{n} = \frac{5-1}{6} = 0{,}667 \quad \Rightarrow \quad \text{mittelstarke positive Korrelation}$$

8.4 Korrelationskoeffizient nach Bravais und Pearson

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

- Maß der Stärke des linearen Zusammenhangs
- Normierung der Kovarianz durch Ausschaltung des Streuungseffekts:
  - Eine Vervielfachung der x- und y- Werte zieht stets eine gleich große Erhöhung der Kovarianz s_{xy} und dem Produkt der beiden Standardabweichungen s_x und s_y nach sich.
  - Nenner und Zähler verändern sich in gleichem Ausmaß, so dass der Wert des Korrelationskoeffizienten unverändert bleibt.
- Der Korrelationskoeffizient ist dimensionslos, da sich die Maßeinheiten, in denen die Merkmalswerte gemessen werden, herauskürzen.
- Wertebereich: $-1 \leq r \leq 1$

Beispiel: Aktien in Euro und Dollar

Die Kovarianz s_{xy} zweier Aktien beträgt 20 €^{2}. Durch eine Umrechnung in Dollar erhöht sich die Kovarianz bei einem Umrechnungskurs von 1,10 $ je 1 € auf:

$$s_{x'y'} = 1{,}10^2 \cdot s_{xy} = 1{,}21 \cdot 20 = 24{,}2[\$^2]$$

Wie verändert sich der Korrelationskoeffizient bei der Währungsumrechnung, wenn als Standardabweichungen s_x = 5 € und s_y = 6 € vorliegen?

- Auf Eurobasis beträgt der Korrelationskoeffizient:

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{20}{5 \cdot 6} = \frac{20}{30} = 0{,}667$$

- Durch die Umrechnung in Dollar werden die beiden Standardabweichungen um den Faktor 1,10 erhöht:

$$s_x = 5 \cdot 1{,}10 = 5{,}5\ [\$]$$

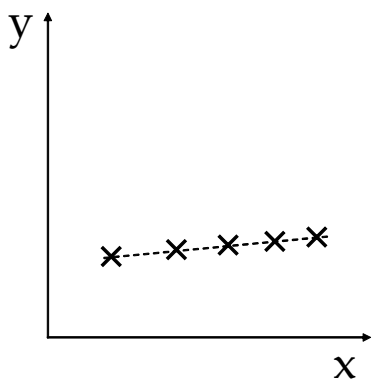
$$s_y = 6 \cdot 1{,}10 = 6{,}6\ [\$]$$

- Da durch eine Lineartransformation Zähler und Nenner um den gleichen Faktor erhöht werden, wirken sich Umrechnungen in andere Einheiten nicht auf die Höhe des Korrelationskoeffizienten aus:

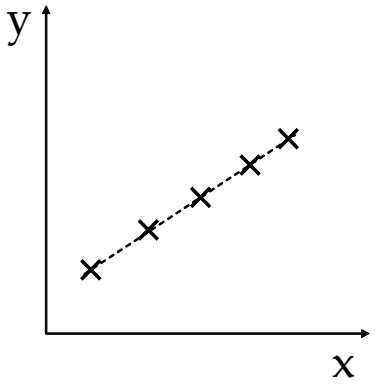
$$r = \frac{s_{x'y'}}{s'_x \cdot s'_y} = \frac{24{,}2}{5{,}5 \cdot 6{,}6} = 0{,}667$$

Perfekte Korrelation und Unkorreliertheit

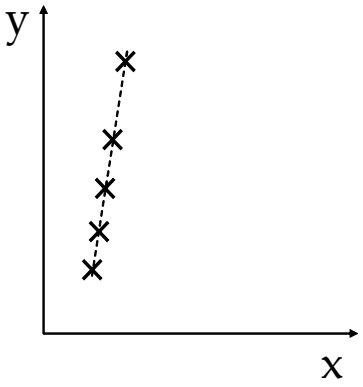
Streuungsdiagramme bei perfekter positiver Korrelation (r = 1):



geringe Steigung

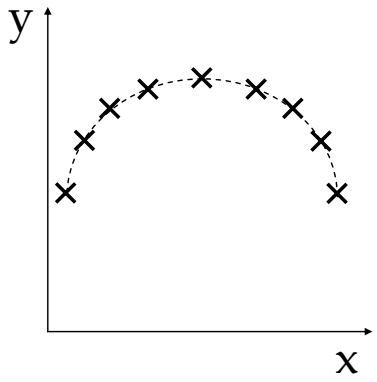


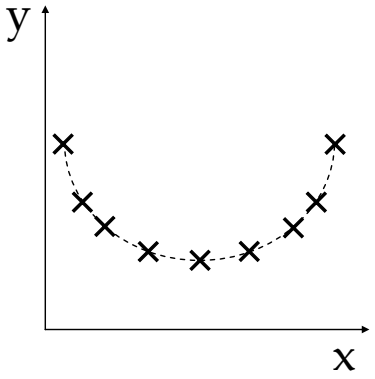
mittlere Steigung

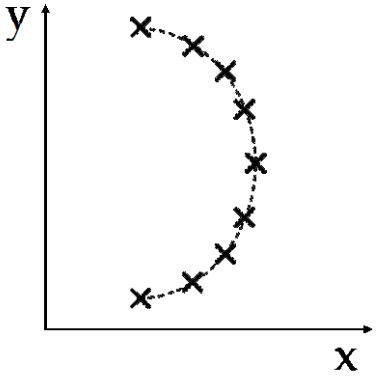


starke Steigung

Parabolische Beziehungen:







Unkorreliertheit (r = 0): Kreisförmige Punktewolke, konstante x-Werte oder y-Werte oder spezielle nichtlineare (z.B. parabolische) Beziehungen

Interpretation des Korrelationskoeffizienten nach Bravais und Pearson

Je näher die Punkte im Streuungsdiagramm an einer steigenden oder fallenden Geraden liegen, desto größer ist der absolute Wert von r.

- $-1 \leq r < -0{,}8$: Starker negativer (linearer) Zusammenhang
- $-0{,}8 \leq r < -0{,}3$: Mittlerer negativer (linearer) Zusammenhang
- $-0{,}3 \leq r < -0{,}1$: Schwacher negativer (linearer) Zusammenhang

- $-0{,}1 \leq r \leq 0{,}1$: Geringer oder kein (linearer) Zusammenhang (Unkorreliertheit)

- $0{,}1 < r \leq 0{,}3$: Schwacher positiver (linearer) Zusammenhang
- $0{,}3 < r \leq 0{,}8$: Mittlerer positiver (linearer) Zusammenhang
- $0{,}8 < r \leq 1$: Starker positiver (linearer) Zusammenhang

Beispiel (Fortsetzung): Werbeausgaben und Umsatz

Für die Varianzen des Beispiels mit Werbeausgaben und Umsatz gilt:

$$s_x^2 = \left(\frac{1}{6} \cdot \sum_{i=1}^{6} x_i^2\right) - \bar{x}^2 = \left[\frac{1}{6} \cdot (3^2 + 5^2 + 4^2 + 7^2 + 8^2 + 10^2)\right] - 6{,}167^2$$

$$= \left(\frac{1}{6} \cdot 263\right) - 38{,}032 = 5{,}801$$

$$s_y^2 = \left(\frac{1}{6} \cdot \sum_{i=1}^{6} y_i^2\right) - \bar{y}^2 = \left[\frac{1}{6} \cdot (1^2 + 4^2 + 2^2 + 4^2 + 5^2 + 6^2)\right] - 3{,}667^2$$

$$= \left(\frac{1}{6} \cdot 98\right) - 13{,}447 = 2{,}886$$

Daraus folgt:

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{3{,}886}{\sqrt{5{,}801} \cdot \sqrt{2{,}886}} = 0{,}950$$

Damit besteht zwischen Umsatz und Werbeausgaben ein starker positiver (linearer) Zusammenhang, d.h. eine starke positive Korrelation.

Berechnung bei unbekannter Kovarianz

Wenn die Kovarianz unbekannt ist, empfiehlt sich zur Berechnung des Korrelationskoeffizienten die Verwendung einer Formel, die von den Summen in der folgenden Arbeitstabelle Gebrauch macht:

i	x_i	y_i	$x_i \cdot yi$	x_i^2	y_i^2
1	x_1	y_1	$x_1 \cdot y_1$	x_1^2	y_1^2
2	x_2	y_2	$x_2 \cdot y_2$	x_2^2	y_2^2
...	...	...	...	...	...
n	x_n	y_n	$x_n \cdot y_n$	x_n^2	y_n^2
$\sum$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i \cdot y_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} y_i^2$

Berechnungsformel:

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

Man erhält diese Berechnungsformel, indem man die mit dem Verschiebungssatz berechneten Varianzen und Kovarianzen verwendet.

Beispiel (Fortsetzung): Werbeausgaben und Umsatz

Die Korrelation zwischen den Werbeausgaben und dem Umsatz lässt sich alternativ auch unter Verwendung der Berechnungsformel ermitteln.

i	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
1	3	1	3·1 = 3	$3^2 = 9$	$1^2 = 1$
2	5	4	5·4 = 20	$5^2 = 25$	$4^2 = 16$
3	4	2	4·2 = 8	$4^2 = 16$	$2^2 = 4$
4	7	4	7·4 = 28	$7^2 = 49$	$4^2 = 16$
5	8	5	8·5 = 40	$8^2 = 64$	$5^2 = 25$
6	10	6	10·6 = 60	$10^2 = 100$	$6^2 = 36$
∑	37	22	159	263	98

$$r = \frac{(6 \cdot 159) - 37 \cdot 22}{\sqrt{[(6 \cdot 263) - (37)^2] \cdot [(6 \cdot 98) - (22)^2]}} = \frac{954 - 814}{\sqrt{109 \cdot 104}} = \frac{140}{147{,}431} = 0{,}950$$

8.5 Rangkorrelationskoeffizient nach Spearman

- Voraussetzung für die Berechnung: **Ordinalskala** (Rangdaten)

Beispiel:

Firmen werden bei einem „Konjunkturtest“ nach ihrer Einschätzung der Produktion, Geschäftslage und den Geschäftsaussichten befragt. Die Merkmalsausprägungen (= Antwortmöglichkeiten) „schlechter“, „gleich“ und „besser“ implizieren ein ordinales Skalenniveau der betrachteten Merkmale.

→ Zur Feststellung der Stärke des Zusammenhangs z.B. zwischen der Geschäftslage und der Produktion aus den Konjunkturtestdaten kann der Pearsonsche Korrelationskoeffizient nicht angewendet werden, da dieser ein metrisches Skalenniveau voraussetzt.

Berechnung des Rangkorrelationskoeffizienten nach Spearman

1. **Schritt**
   - Transformation der Ursprungsdaten von X und Y in Rangdaten mit Rang von x_i: $r(xi)$ und Rang von y_i: $r(y_i)$.
   - Falls Bindungen vorliegen, d.h. zwei oder mehrere Merkmalsträger denselben Merkmalswert haben, wird ihnen jeweils das arithmetische Mittel der zu vergebenden Ränge zugeordnet.
2. **Schritt:** Bildung von Rangdifferenzen: $d_i = r(x_i) - r(y_i)$
3. **Schritt:**
   - Berechnung des Rangkorrelationskoeffizienten

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)}$$

   - Wertebereich: $-1 \leq r_S \leq 1$
   - Obere Schranke ($r_S = 1$): In diesem Fall gilt, dass aus $r(x_i) < r(x_j)$ stets $r(y_i) < r(y_j)$ folgt (Übereinstimmung der Ränge bei den Merkmalen X und Y).
   - Untere Schranke ($r_S = -1$): In diesem Fall gilt, dass aus $r(x_i) < r(x_j)$ stets $r(y_i) > r(y_j)$ folgt (perfekte inverse Rangfolge Y im Vergleich zu X).

Beispiel: Produkttest (I)

Acht Personen sollen im Rahmen eines „Produkttests“ die Qualität (X) und den Preis (Y) einer Ware auf einer Schulnotenskala (1, 2, 3, 4, 5, 6) beurteilen.

Alle verbundenen Beobachtungen gehen aus folgender Tabelle hervor:

i	x_i (Qualitätsbeurteilung)	y_i (Preisbeurteilung)
1	1	2
2	2	3
3	3	5
4	2	3
5	1	1
6	4	4
7	2	3
8	5	6

Beispiel: Produkttest (II)

1. **Schritt:** Rangfolge bilden

2. **Schritt:** Rangdifferenzen bilden

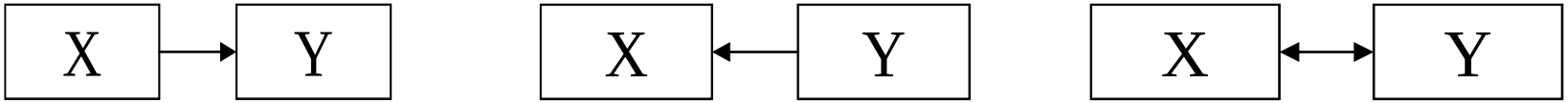
i	x_i	y_i	$r(x_i)$	$r(y_i)$	$d_i = r(x_i) - r(y_i)$	d_i^2
1	1	2	1,5	2	-0,5	0,25
2	2	3	4	4	0	0
3	3	5	6	7	-1	1
4	2	3	4	4	0	0
5	1	1	1,5	1	0,5	0,25
6	4	4	7	6	1	1
7	2	3	4	4	0	0
8	5	6	8	8	0	0
∑						2,5

3. **Schritt:** Berechnung des Rangkorrelationskoeffizienten

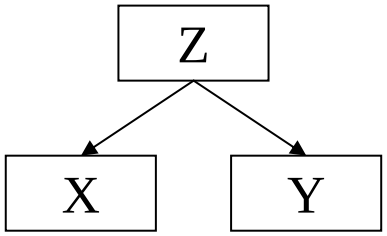
$$r_S = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 2{,}5}{8 \cdot (8^2 - 1)} = 1 - \frac{15}{504} = 0{,}970$$

8.6 Kausalität und Scheinkorrelation

- Unter **Kausalität** (= ursächlicher Zusammenhang) versteht man eine einseitige oder wechselseitige Abhängigkeit zwischen den Merkmalen X und Y.
- Mögliche Kausalitäten:



- Bei absolut hoher Korrelation: Kausaler Zusammenhang oder Scheinkorrelation
- Eine hohe **Scheinkorrelation** liegt vor, wenn eine absolut hohe Korrelation ermittelt wird, ohne dass die beiden betrachteten Merkmale in einem kausalen Zusammenhang stehen.
- Oft wird die Korrelation zwischen den betrachteten Merkmalen X und Y durch eine dritte Größe Z hervorgerufen, die sowohl auf X als auch auf Y einwirkt:



Beispiel: Sonnenbrände und Mineralwasser

- Betrachtet man die Zahl der Sonnenbrände (X) und den Verbrauch von Mineralwasser (Y), dann wird man in den Sommermonaten möglicherweise einen hohen Korrelationskoeffizienten ermitteln können.
- Dieser hohe Korrelationskoeffizient drückt jedoch keine kausale, d.h. ursächliche Beziehung zwischen X und Y aus.
- Vielmehr lässt er sich auf die Zahl der Hitzetage (= eine dritte Größe Z) zurückführen, die sowohl die Zahl der Sonnenbrände (X) als auch den Verbrauch von Mineralwasser (Y) ansteigen lässt.
- Die hohe Korrelation zwischen X und Y ist daher eine Scheinkorrelation.

In den Wirtschaftswissenschaften kommt das Phänomen der Scheinkorrelation oft bei der Analyse von Zeitreihen vor.

→ Eine Scheinkorrelation kann in einem Beobachtungszeitraum durch Trends in den Variablen zustande kommen.