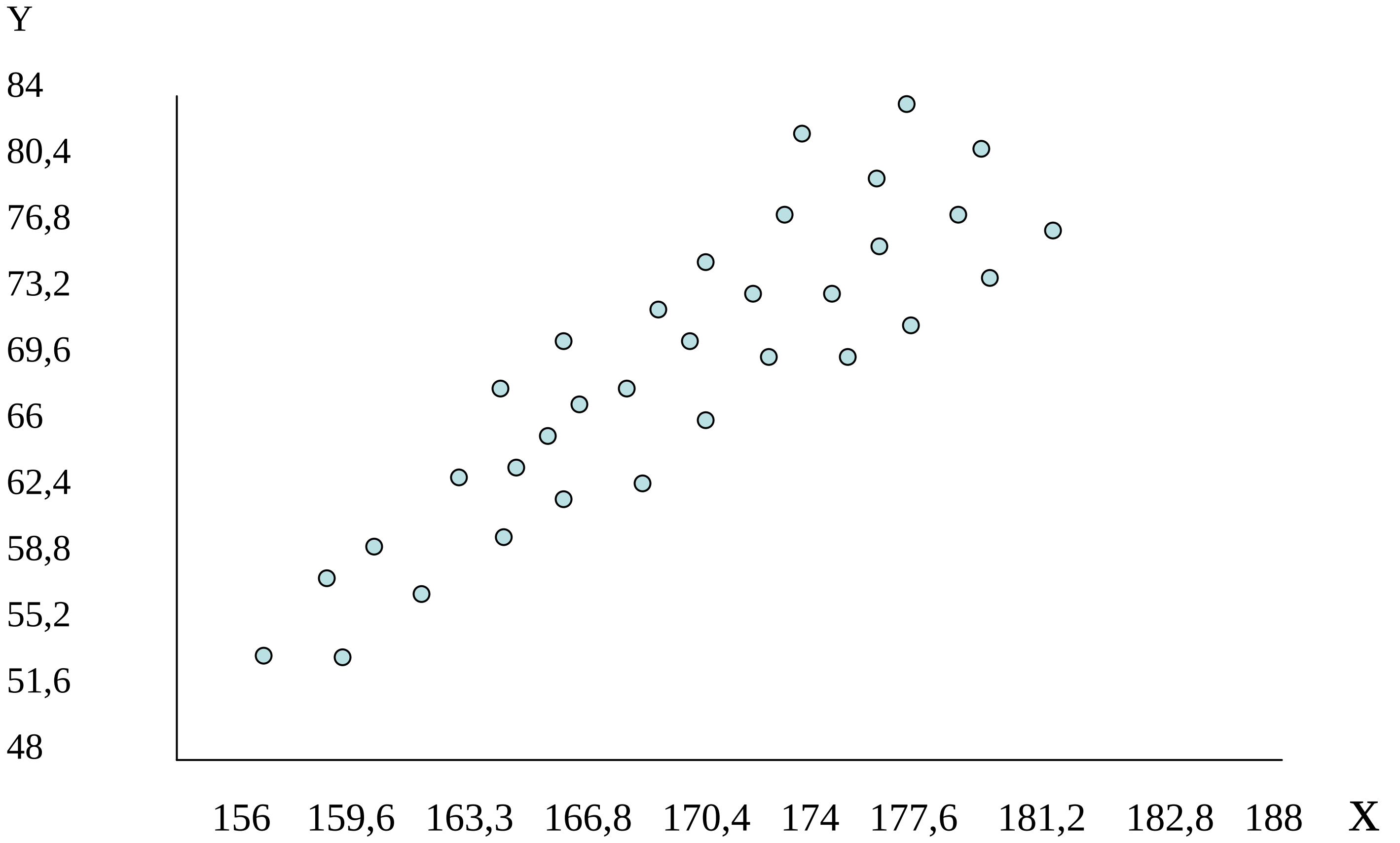
Streudiagramm

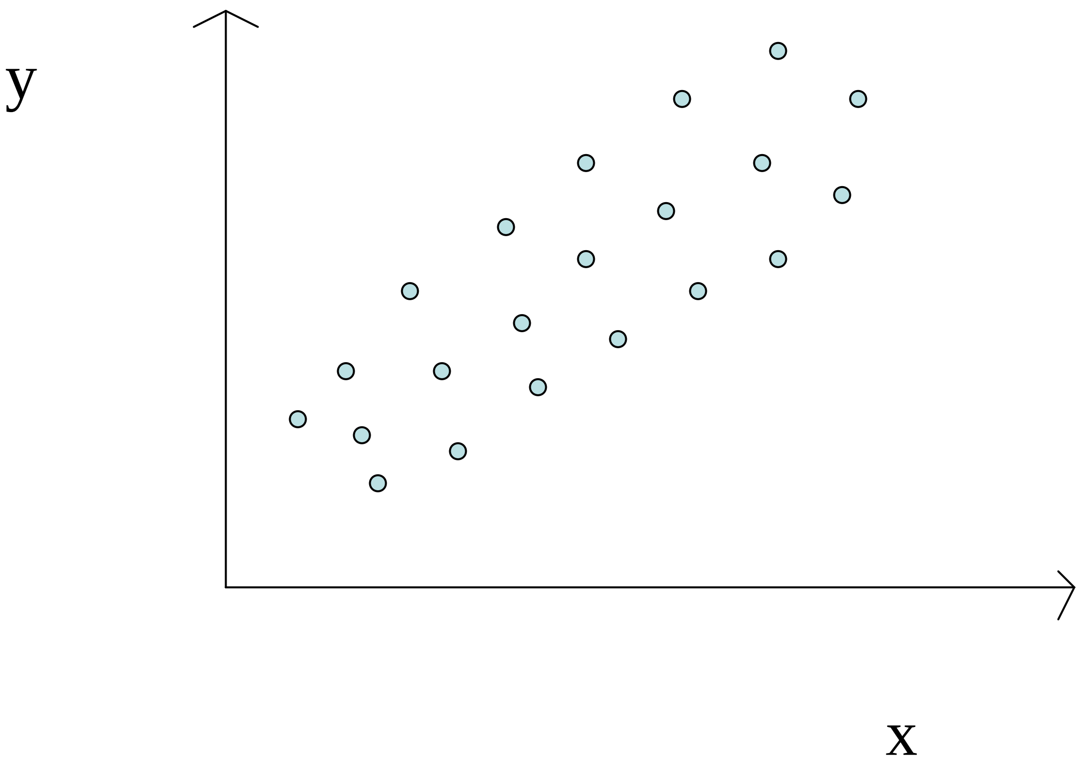
Beispiel: Die Beziehung zwischen Körpergröße und Gewicht bei einer Studentengruppe im Kurs Statistik



Y-Achse -> vertikal ≙ Körpergewicht in kg

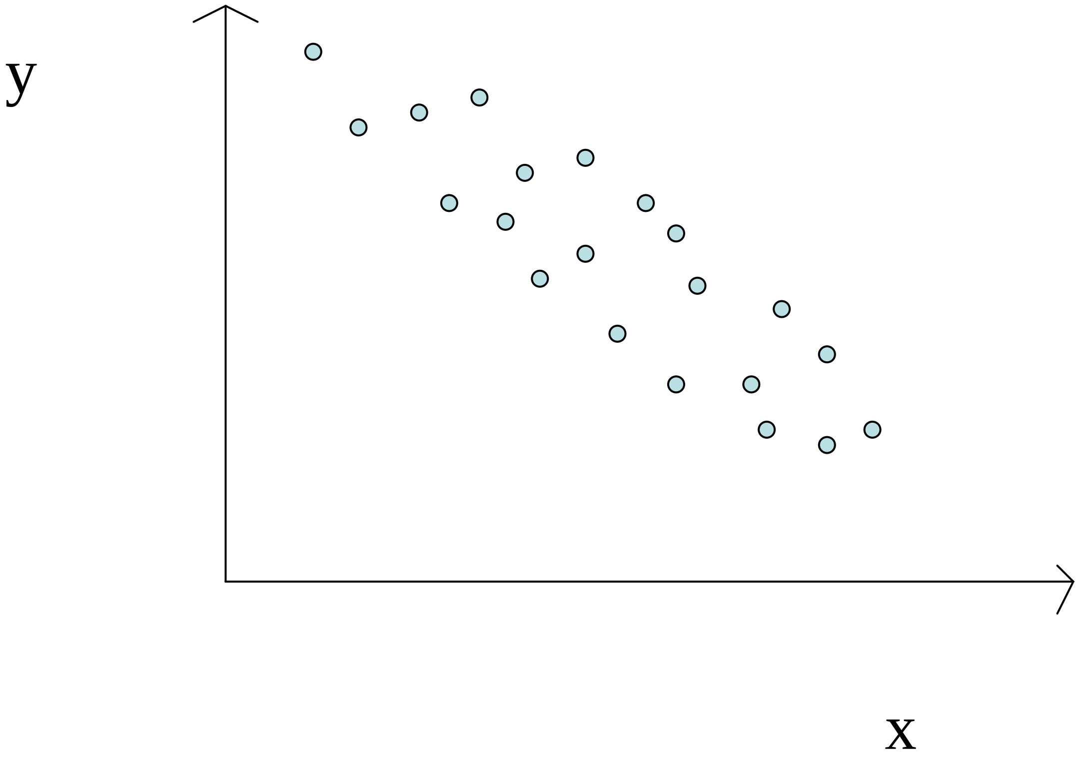
X-Achse -> horizontal ≙ Körpergröße in cm

Richtung der Beziehung: Positive Korrelation



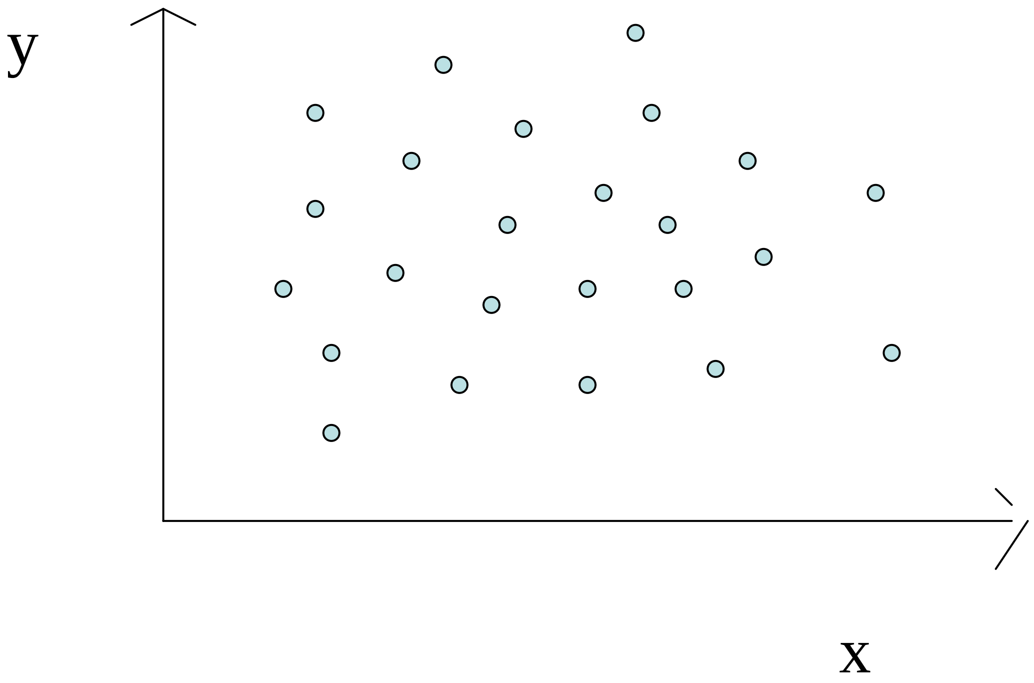
Es besteht ein positiver linearer Zusammenhang zwischen den Variablen; wenn die eine Variable ansteigt, dann erhöht sich auch die andere Variable.

Richtung der Beziehung: Negative Korrelation



Es besteht ein negativer linearer Zusammenhang zwischen den Variablen; eine Erhöhung des Ausprägungsgrades der einen Variablen verursacht ein Absinken des Ausprägungsgrades der anderen Variable.

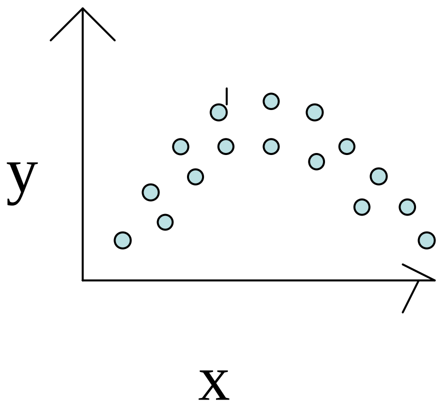
Nullkorrelation

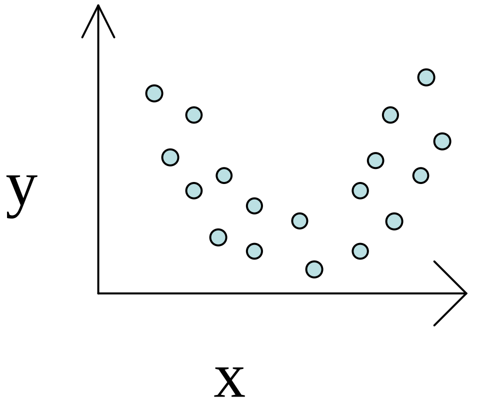


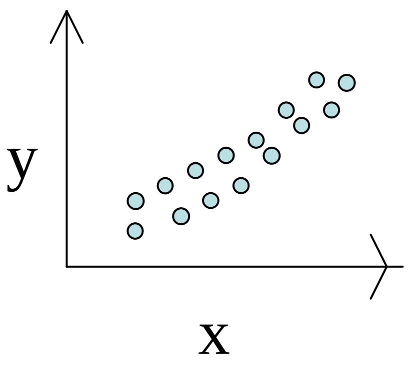
Es besteht kein Zusammenhang zwischen den Variablen.

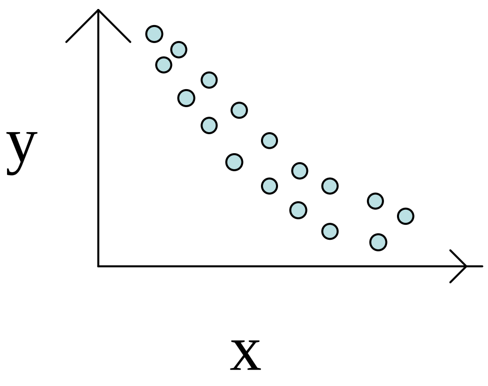
Kurvenlineare Beziehung

Es besteht ein Zusammenhang zwischen den Variablen, der jedoch nicht linear ist.









Pearsonsche Produkt-Moment-Korrelationskoeffizient r

Die Zahlenwerte des Produkt-Moment-Korrelationskoeffizienten variieren zwischen –1 und +1:

0 ≙ Keine Beziehung

-1 ≙ perfekte negative Beziehung

+1≙ perfekte positive Beziehung

Grundsatz: r kann nur bei metrischen Variablen verwendet werden und auch nur dann, wenn eine ***lineare Beziehung*** zwischen den Variablen besteht.

Berechnung von r

1.Formel

Den Korrelationskoeffizienten r erhalten wir, indem die Kovarianz zweier Variablen durch das Produkt der Standardabweichungen der Variablen dividiert wird.

$$r = \frac{\operatorname{cov}(x,y)}{s_x \bullet s_y} = \frac{\dfrac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n}}{\sqrt{\dfrac{\sum (x_i - \overline{x})^2}{n}} \bullet \sqrt{\dfrac{\sum (y_i - \overline{y})^2}{n}}}$$

verkürzt:

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \bullet \sum (y_i - \overline{y})^2}}$$

Beispiel zur Berechnung von r

Messwertpaare (x;y) : (9;7), (7;8), (8;7), (6;5), (8;9), (7;6)

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
	9	7	1,5	0	2,25	0	0
	7	8	-0,5	1	0,25	1	-0,5
	8	7	0,5	0	0,25	0	0
	6	5	-1,5	-2	2,25	4	3
	8	9	0,5	2	0,25	4	1
	7	6	-0,5	-1	0,25	1	-0,5
$\sum$	45	42	0	0	5,5	10	4

$$\bar{x} = 7,5$$

$$\bar{y} = 7$$

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \bullet \sum (y_i - \overline{y})^2}}$$

$$r = \frac{4}{\sqrt{5{,}5 \bullet 10}} = 0{,}54$$

[Da diese Formel auf der Berechnung von Mittelwerten basiert, bereitet sie lästige Rechenarbeit, wenn die Mittelwerte Zahlen mit mehreren Dezimalstellen sind.]

2.Formel

$$r = \frac{n \bullet \sum x \bullet y - \sum x \sum y}{\sqrt{(n \bullet \sum x^2 - (\sum x)^2)(n \bullet \sum y^2 - (\sum y)^2)}}$$

x	y	$x \bullet y$	x^2	y^2
9	7	63	81	49
7	8	56	49	64
8	7	56	64	49
6	5	30	36	25
8	9	72	64	81
7	6	42	49	36
Σ 45	42	319	343	304

n=6

$$r = \frac{6 \bullet 319 - 45 \bullet 42}{(6 \bullet 343 - 45^2)(6 \bullet 304 - 42^2)}$$

$$r = \frac{24}{\sqrt{33 \bullet 60} = \dfrac{24}{44{,}49}} = 0{,}54$$

Die Bedeutsamkeit von r oder Die Stärke der Beziehung

bis 0,2	≙	schwach
0,2 bis 0,4	≙	niedrig
0,4 bis 0,7	≙	mäßig
0,7 bis 0,9	≙	hoch
Über 0,9	≙	sehr hoch