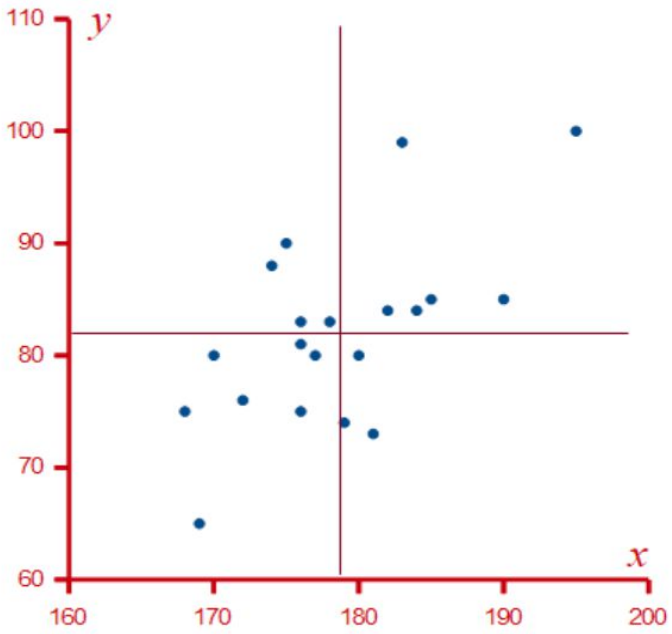
Calcul du coefficient de corrélation linéaire

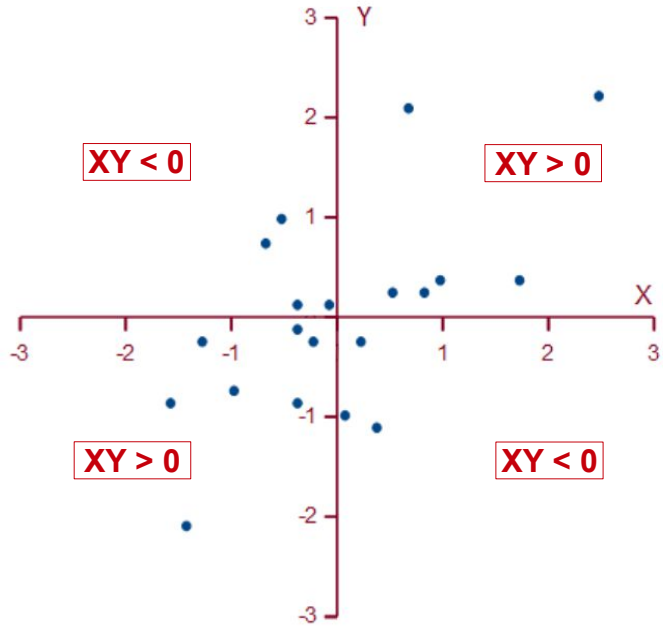
Le principe du calcul est résumé dans le second graphique ci-dessous.

Les variables x et y qui peuvent par exemple représenter des tailles et des masses, sont remplacées par les variables X et Y qui sont les écarts à la moyenne divisées par l'écart-type.

On obtient ainsi deux variables ***centrées-réduites***, leur moyenne est nulle et leur écart-type est égal à 1.

Cela permet notamment de standardiser les dispersions qui sont fortement influencées par le choix des unités.





Les quatre régions obtenues sont caractérisées par le signe du produit XY.

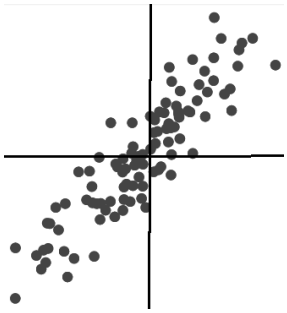
Le coefficient de corrélation linéaire est la moyenne de tous les produits XY.

$$r = \frac{1}{n}\sum_{i=1}^{n} XY = 0{,}62$$

C'est un nombre compris entre -1 et 1

Il est d'autant plus élevé que les régions de la première diagonale (XY > 0) sont remplies.

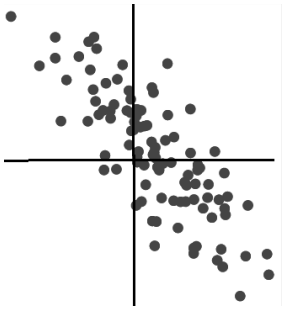
Exemples :



$r = 0{,}9$



$r = -0{,}4$



$r = -0{,}8$

Si le coefficient est égal à 1, le nuage est réduit à une droite d'où le nom de corrélation linéaire.

FORMULES

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

$$X = \frac{x - \bar{x}}{\sigma_x} \qquad \text{Moyenne : } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{Écart-type : } \sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

CALCUL AU TABLEUR

i	x_i	y_i	X_i	Y_i	$X_i Y_i$
1	180	80	0,23	0,25	-0,06
2	175	90	-0,53	0,98	-0,52
3	170	80	-1,28	-0,25	0,31
4	190	85	1,73	0,37	0,64
5	168	75	-1,58	-0,86	1,36
6	176	83	-0,38	0,12	-0,05
7	182	84	0,53	0,25	0,13
8	195	100	2,48	2,21	5,49
9	178	83	-0,08	0,12	-0,01
10	181	73	0,38	-1,11	-0,42
11	172	76	-0,98	-0,74	0,72
12	177	80	-0,23	-0,25	0,06
13	184	84	0,83	0,25	0,20
14	174	88	-0,68	0,74	-0,50
15	179	74	0,08	-0,98	-0,07
16	183	99	0,68	2,09	1,41
17	169	65	-1,43	-2,09	2,98
18	185	85	0,98	0,37	0,36
19	176	81	-0,38	-0,12	0,05
20	176	75	-0,38	-0,86	0,32
Moyennes	178,5	82	0	0	**0,62**
Écarts-types	6,7	8,1	1	1	