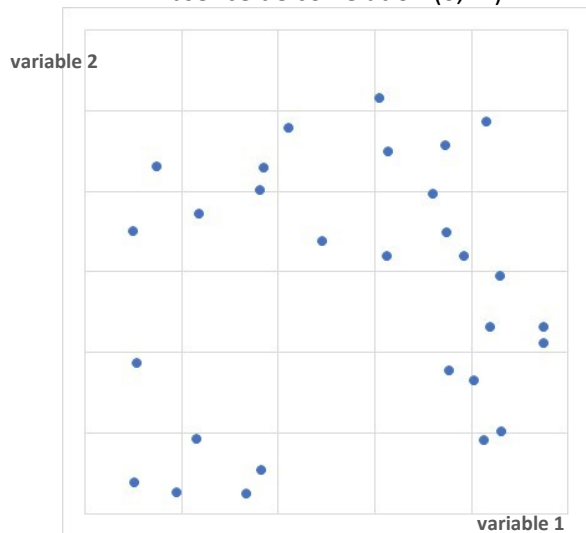


Définition simplifiée de la corrélation

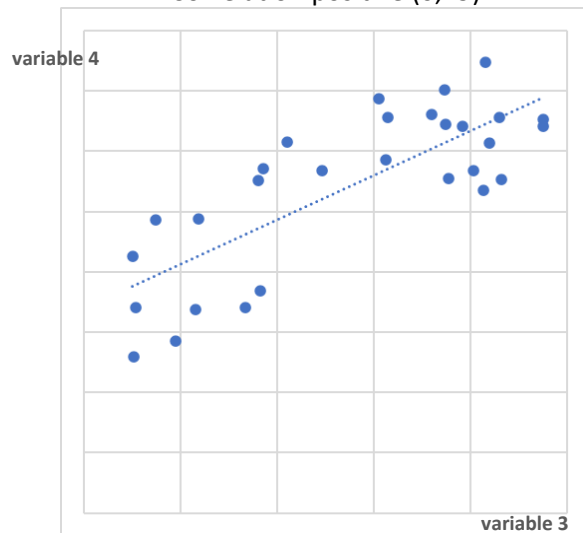
La corrélation est l'expression de l'intensité de la relation entre deux variables. Par exemple, il est possible qu'il y ait une forte corrélation entre un temps chaud et ensoleillé et le nombre de glaces vendues. Si les deux variables vont dans la même direction, c'est la corrélation positive. À l'inverse, si elles vont dans des directions opposées, c'est la corrélation négative. Autrement formulé, dans le cas des corrélations de deux variables, les variations des valeurs de l'une sont concomitantes des variations des valeurs de l'autres.

Selon la nature des variables dont on analyse la corrélation, différents indicateurs statistiques peuvent être mobilisés. Dans le cas le plus classique de deux variables « quantité », on utilise habituellement le coefficient de corrélation linéaire qui permet de mesurer à la fois la force et le sens d'une association. Variant de -1 à +1, il vaut 0 lorsqu'il n'existe pas d'association. Plus ce coefficient est proche de -1 ou +1, plus l'association entre les deux variables est forte, jusqu'à être parfaite.

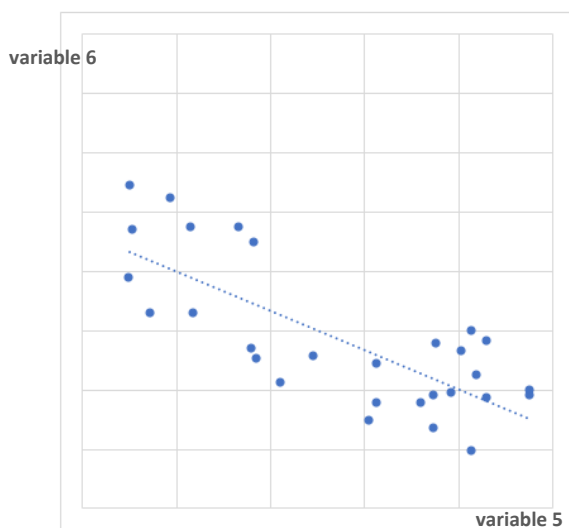
Absence de corrélation (0,12)



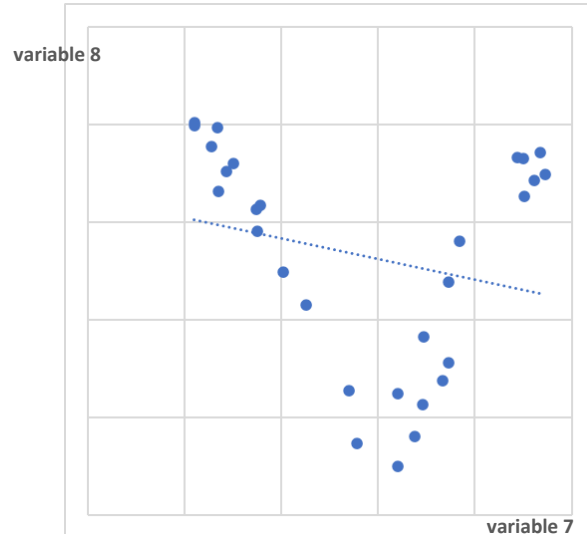
Corrélation positive (0,79)



Corrélation négative (-0,76)



Fausse « faible corrélation négative » (-0,23)
relation non linéaire



Ces illustrations reposent sur des phénomènes continus (dans notre premier exemple la température et le nombre de glace vendues). L'indicateur de corrélation retenu est alors le Coefficient de corrélation linéaire de Pearson.

La notion de corrélation peut être également mobilisée pour des variables de type catégories (et non pas quantité) : le genre, la région, le niveau de formation, ... Dans tous les cas, on cherche toujours à analyser si le fait de changer de catégorie au sein d'une variable est se traduit également par le fait de changer de catégorie pour l'autre variable. Pour ces variables, également appelées variables qualitatives, de nombreux autres indicateurs permettent de mesurer la corrélation, parmi lesquelles le χ^2 , le V de Cramer, ...

Elle permet également de mesurer les liaisons entre une variable qualitative et un phénomène quantitatif au travers d'indicateurs tels le η^2 .

Enfin, il est important de préciser qu'observer une corrélation n'induit pas forcément une causalité. La littérature regorge d'exemples illustrant ce propos. Ainsi, par exemple, on note que dans les communes abritant des cigognes, le taux de natalité est plus élevé que dans l'ensemble du pays. Par boutade, on pourrait en conclure que les cigognes apportent les bébés. Une explication plus rationnelle repose sur le fait que les cigognes nichent de préférence dans les villages plutôt que dans les grandes agglomérations, or il se trouve que la natalité est plus forte en milieu rural que dans les villes.

Cette note reprend très largement les articles en ligne suivant :

[https://fr.wikipedia.org/wiki/Corr%C3%A9lation_\(statistiques\)](https://fr.wikipedia.org/wiki/Corr%C3%A9lation_(statistiques))

<https://soepidemio.com/2016/11/14/correlation-statistique-prudence-a-linterpretation/>

https://www.lemonde.fr/les-decodeurs/article/2019/01/02/correlation-ou-causalite-brillez-en-societe-avec-notre-generateur-aleatoire-de-comparaisons-absurdes_5404286_4355770.html

<https://cortecs.org/la-zetetique/effets-cigogne-correlation-vs-causalite/>