

Corrélation et causalité

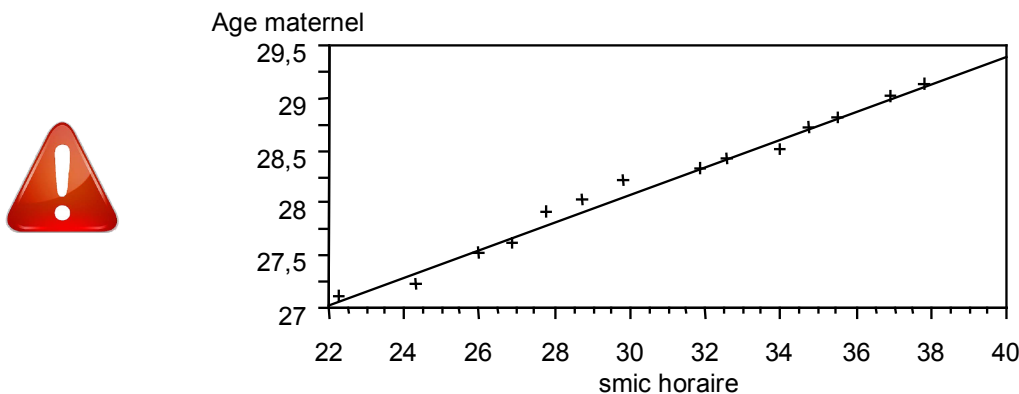
-Une maxime statistique est “ ***une corrélation forte n’implique pas nécessairement une causalité*** ” ; ce qui s’énonce aussi :

Si un nuage de points est presque rectiligne, cela n’implique pas nécessairement de relation de cause à effet entre le phénomène mesuré par les abscisses et celui qui est mesuré par les ordonnées ;

Si cette maxime ne relève que d’un argument d’autorité, et n’est justifiée que par l’étude d’exemples choisis pour leur caractère absurde, elle risque d’être irrésistiblement récusée au premier graphique marquant (*tel l’alignement de 10 points, dont l’abscisse donne le nombre d’entrées réalisées par les films étiquetés “ film violent ” ces dix dernières années et l’ordonnée le nombre d’agressions sur la voie publique pendant les mêmes années*). Il est important de comprendre que :

- si l’abscisse et l’ordonnée des points sont presque des fonctions affines d’une même variable t (le temps par exemple), alors “ mathématiquement ” le nuage des points sera presque une droite ; les points n’ont pas une “ tendance naturelle ” à s’aligner, et un alignement appelle une explication : celle-ci est le plus souvent à rechercher dans l’existence d’une dépendance linéaire à un même facteur (appelé en médecine facteur de confusion).

- l’évolution de très nombreux phénomènes est en première approximation affine (même les phénomènes exponentiels à taux d’accroissements faibles en sont des cas particuliers).



Dans le graphique ci-dessus, chaque point correspond à une année, (De gauche à droite : de 1983 à 1996) ; l’abscisse est la valeur du SMIC horaire en début d’année et l’ordonnée l’âge moyen des femmes ayant eu un bébé cette année-là.

Le coefficient de corrélation linéaire pour ces 14 points est 0,99.
Cet exemple illustre simplement qu’en première approximation, entre 1983 et 1996, l’accroissement absolu annuel du SMIC d’une part, et de l’âge maternel d’autre part, sont approximativement constants.