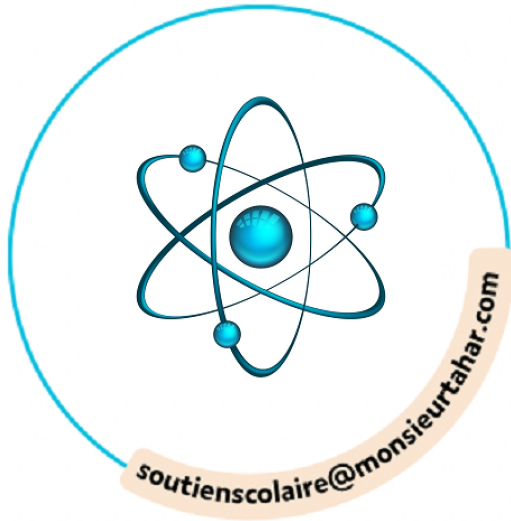
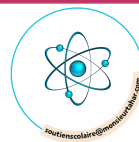


SES



FICHE METHODE

Distinguer Corrélation et Causalité



« Faut-il manger du chocolat pour avoir des prix Nobel ? » C'était le titre d'un article¹ relatant les résultats d'une étude montrant que les pays où les habitants mangent le plus de chocolat obtiennent plus de prix Nobel. Ce type de constat s'appelle une **corrélation**, mais ne dit en rien que manger du chocolat augmente l'intelligence !

1. « Faut-t-il manger du chocolat pour avoir des prix Nobel ? », *La Dépêche*, 12 octobre 2012.

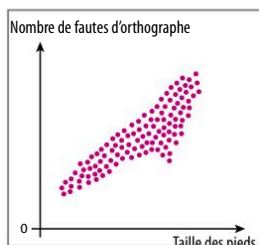
Qu'est-ce qu'une corrélation ?

Une **corrélation** est un constat statistique que deux variables évoluent dans le même sens (corrélation positive) ou en sens inverse (corrélation négative). Quand les évolutions des deux variables paraissent indépendantes l'une de l'autre, il n'y a pas de corrélation.

Imaginons que vous lanciez une étude statistique auprès de l'ensemble des élèves d'une école primaire. Vous faites passer à tous les élèves un test d'orthographe sous la forme d'une courte dictée, comptez le nombre de fautes par élève... et mesurez aussi la taille de leurs pieds. Voici les trois cas de figure possibles pour les résultats.

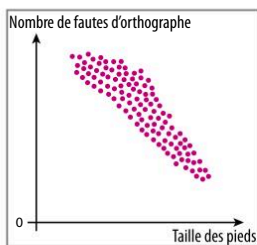
EXEMPLE

Cas 1 : Les deux variables évoluent dans le même sens.



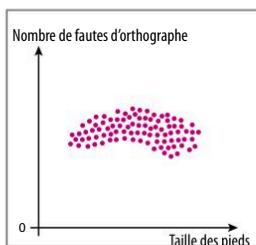
Plus les élèves ont de grands pieds, plus ils font de fautes d'orthographe → Il s'agit d'une **corrélation positive**.

Cas 2 : Les deux variables évoluent en sens inverse



Plus les élèves ont de grands pieds, moins ils font de fautes d'orthographe → Il s'agit d'une **corrélation négative**.

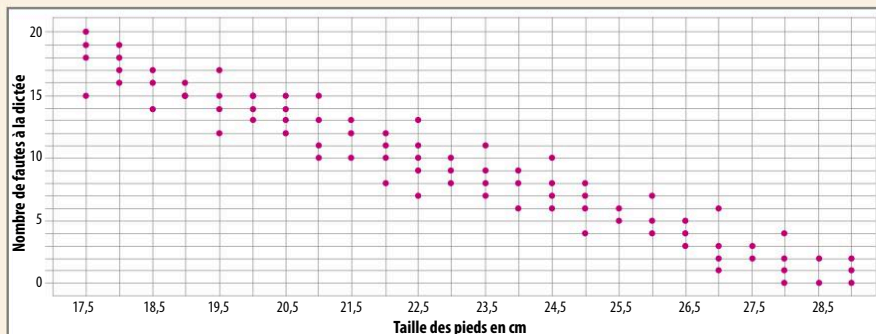
Cas 3 : Les deux variables évoluent indépendamment l'une de l'autre



Quand les élèves ont de plus grands pieds, ils font parfois plus, parfois moins de fautes d'orthographe → Il n'y a **pas de corrélation**.

De la corrélation à la causalité

Voici les résultats trouvés dans une enquête :



→ Ici, il y a une corrélation, et elle est négative : quand la taille des pieds des élèves est plus élevée, ils font moins de fautes.

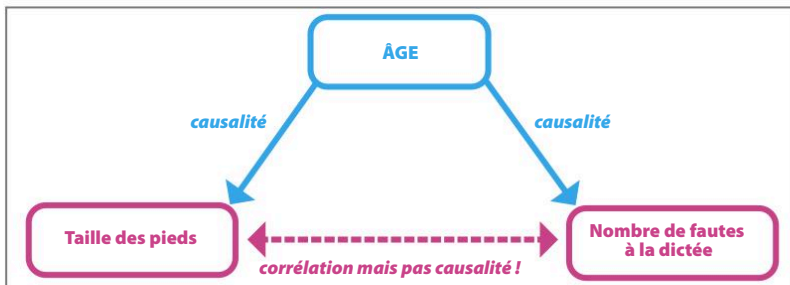


Pour dégager une **causalité** entre ces deux phénomènes mesurés par les variables A (taille des pieds) et B (nombre de fautes), vous devez élaborer une relation de cause à effet sous la forme d'une déduction : A entraîne B ou B entraîne A.

- Ici, iriez-vous imaginer que les pieds plus grands permettent de développer les apprentissages en orthographe ?
- Ou que les apprentissages en orthographe accélèrent la croissance des pieds ?

Cet exemple montre qu'une **corrélation n'implique pas nécessairement de lien de causalité entre deux variables A et B**.

Une des raisons peut être qu'il existe une troisième variable, C, qui explique la corrélation trouvée entre A et B. Ici, c'est l'âge des enfants de cette école : entre 6 et 11 ans, c'est l'âge qui explique les progrès en orthographe et la croissance des pieds !



Exemple inspiré de Nicolas GAUVRIT, *Statistiques, Méfiez-vous !*, Ellipses, 2007.

Conclusion

Une **corrélation n'est donc qu'un indice d'une causalité possible** entre deux variables, mais un **indice à considérer avec précaution**, en retirant l'influence de variables cachées.

- Ici, en ne prenant que les élèves de la même classe et du même âge, il n'y a plus de corrélation entre la taille des pieds et le nombre de fautes d'orthographe !

Une **corrélation peut aussi être le simple fruit du hasard**. Vous pouvez vous amuser à générer des corrélations absurdes, sans liens de causalité, avec cette animation : 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

- Par exemple, depuis vingt ans, il y a une excellente corrélation positive entre deux variables qui baissent régulièrement : les ventes de cigarettes et le nombre d'homicides en France.

À RETENIR

- Une **corrélation** est un lien statistique entre deux variables.
- Une **corrélation** peut être **positive** ou **négative**.
- Pour dégager une **causalité**, il faut élaborer une **relation de cause à effet** sous la forme d'une déduction : A entraîne B.
- Une **corrélation n'est pas toujours synonyme de causalité** : une **variable cachée** peut expliquer le lien statistique ou la corrélation peut être un simple hasard.