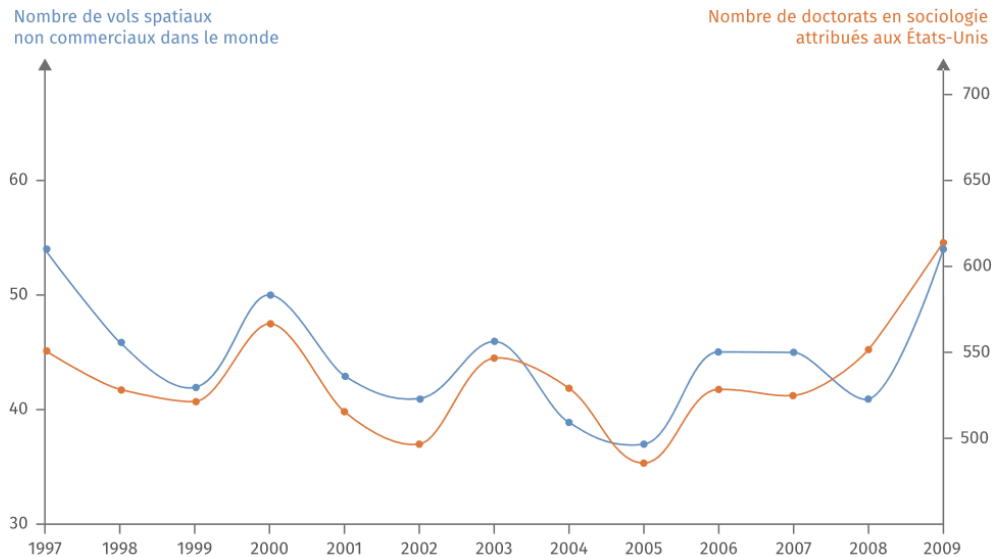


CORRIGÉ - TD - Maîtriser les notions de corrélation et de causalité

En partant d'une question incongrue, ce TD vise à montrer que deux phénomènes corrélés peuvent entretenir un lien de causalité mais que ce n'est pas toujours le cas. La distinction entre corrélation et causalité est à la base de la compréhension des phénomènes étudiés en SES. Elle permet aussi d'exercer son esprit critique en décelant des situations où la confusion règne parfois.

Document 1 - Un lien entre les lancements spatiaux et les doctorats de sociologie ?



Étape 1. Comprendre la signification des données.

1. Faites des phrases avec les données de 2009.

Le nombre de lancements spatiaux se lit sur l'axe de gauche, celui de doctorats en sociologie se lit à droite.

En 2009, il y a eu 53 vols spatiaux non commerciaux dans le monde et 610 doctorats en sociologie décernés aux États-Unis.

Étape 2. Observer l'évolution des données.

2. Montrez qu'il y a une corrélation entre l'évolution du nombre de vols spatiaux et l'évolution du nombre de doctorats en sociologie décernés aux États-Unis.

Dans un premier temps, il faut étudier l'évolution du nombre de vols spatiaux, puis celle du nombre de doctorats et enfin voir si on peut établir un lien entre les deux.

Lorsque la variable A (vols spatiaux non commerciaux) augmente, la variable B (nombre de doctorats en sociologie décernés aux USA) augmente également. Il y a donc une corrélation entre les deux variables A et B.

Étape 3. S'interroger sur un possible lien de causalité.

3. Cette corrélation montre-t-elle une relation de causalité entre ces deux données ?

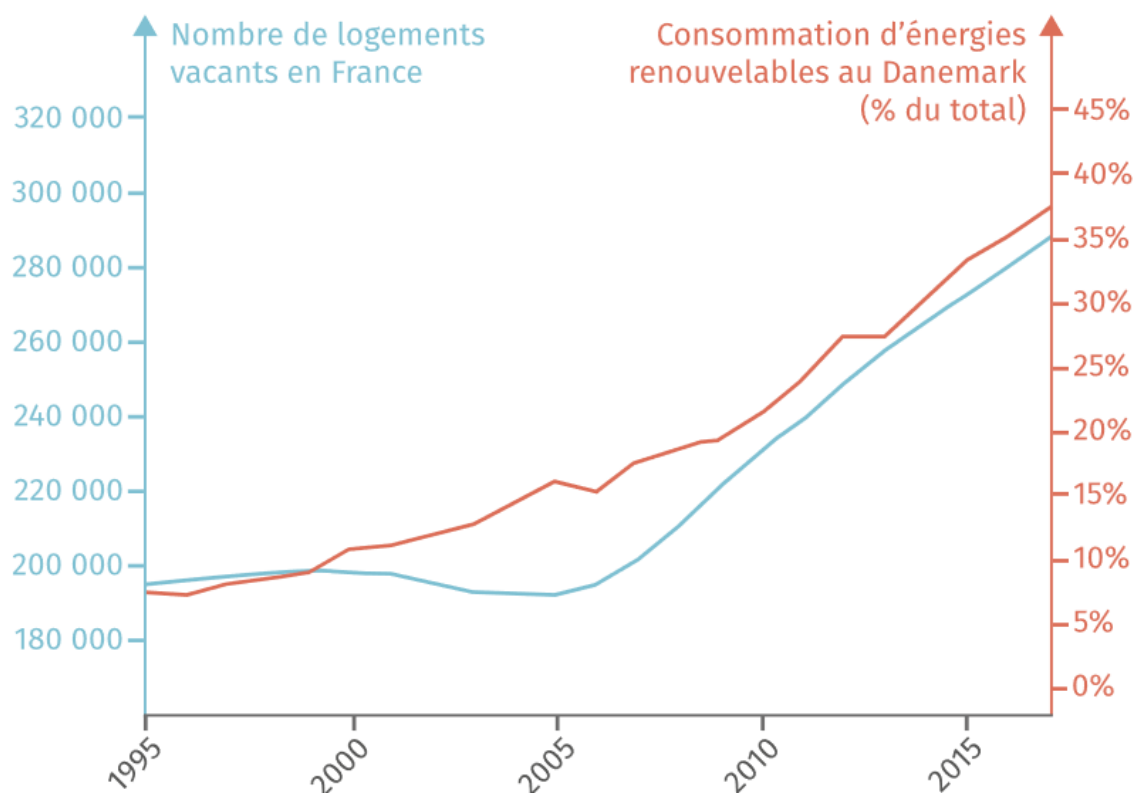
Une fois la corrélation établie entre deux variables A et B, un lien de causalité se cherche en se demandant si A peut expliquer B ou B peut expliquer A. Si ce lien n'existe pas, c'est qu'il n'y a pas de relation de causalité entre ces deux variables.

On ne peut pas dire que A est la cause de B. Le lien n'existe pas, il y a donc corrélation mais pas causalité. Ni A ni B ne sont la conséquence de l'autre.

Un 3^{ème} phénomène, C, le **facteur confondant**, peut-il être la cause à la fois de A et de B ?

Oui la richesse d'un pays, plus un pays est riche, plus il pourra consacrer d'argent à l'éducation (doctorats) et à la recherche spatiale.

Document 2 - Un lien entre les logements vacants et la consommation d'énergies renouvelables



1. Faites des phrases présentant les données pour 2010.

En 2010, il y avait 220 000 logements vacants en France et le Danemark utilisait 20 % d'énergie renouvelable.

2. Montrez qu'il y a une corrélation entre le nombre des logements vacants en France et la consommation d'énergies renouvelables au Danemark.

Lorsque la variable A (logements vacants en France) augmente, la variable B (Part d'énergies renouvelables utilisée au Danemark) augmente également. Il y a donc une corrélation entre les deux variables A et B.

3. Cette corrélation montre-t-elle une relation de causalité entre ces deux variables ?

On ne peut pas dire que A est la cause de B. Le lien n'existe pas, il y a donc corrélation mais pas causalité.

Ni A ni B ne sont la conséquence de l'autre.

Profils sociologiques des électeurs au premier tour de la présidentielle de 2017

Les données sont exprimées en pourcentage. Le total de chaque ligne ne fait pas 100 % car certains candidats ont été retirés.

		J.-L. Mélenchon	B. Hamon	E. Macron	F. Fillon	M. Le Pen
Niveau de diplôme	Inférieur au baccalauréat	17	4	19	19	30
	Niveau baccalauréat	21	6	24	15	24
	Supérieur à bac +3	20	10	30	24	9
Âge	18-24 ans	30	10	18	9	21
	25-34 ans	24	8	28	8	24
	35-49 ans	22	7	21	11	29
	50-59 ans	21	6	21	13	27
	60-69 ans	15	5	26	27	19
	70 ans et plus	9	3	27	45	10

1. Y a-t-il une corrélation entre le niveau de diplôme et la fréquence du vote pour Emmanuel Macron ? Y a-t-il une corrélation entre le niveau de diplôme et la fréquence du vote pour Marine Le Pen ?

Réponse = affirmation :

Dans les deux cas il y a des **corrélations**.

Explication :

Plus le niveau de diplôme augmente, plus la proportion d'électeurs d'Emmanuel Macron augmente (**corrélation positive**). À l'inverse, la faiblesse du niveau de diplôme accroît la tendance/propension à voter pour Marine Le Pen (**corrélation négative**).

Illustration :

Par exemple, 30 % des électeurs de Marine Le Pen avaient un diplôme inférieur au baccalauréat alors que 9 % de ses électeurs avaient un diplôme supérieur à bac +3.

2. Pour quels candidats observe-t-on une corrélation entre l'âge et la fréquence du vote ? Selon vous, peut-il y avoir une relation de causalité entre l'âge et le vote ?

1^{ère} question

Réponse = affirmation :

On observe de nettes corrélations entre l'âge et la fréquence du vote pour Jean-Luc Mélenchon, Benoît Hamon et François Fillon.

Explication :

En effet, pour les deux premiers candidats, la proportion d'électeurs décroît avec l'âge, alors que pour François Fillon, la proportion d'électeurs augmente en même temps que l'âge.

Illustration :

Par exemple, 30 % des 18-24 ans ont voté pour Jean-Luc Mélenchon et 9 % des 70 ans et plus ont voté pour lui. À l'inverse, 9 % des 18-24 ans ont voté pour François Fillon et 45 % des 70 ans et plus ont voté pour lui.

2^{ème} question

Réponse = Affirmation

Il peut y avoir une relation de causalité entre l'âge et le vote.

Explication :

Certaines classes d'âge et certaines générations adhèrent plus ou moins aux programmes des différents candidats. D'autre part, l'âge joue sur la situation socio-économique des individus et peut orienter leur vote.

Cependant

Ce n'est pas l'âge qui détermine le vote, mais les valeurs ou la situation socio-économique des individus liée à leur âge et à leur génération qui influence le vote.

À l'aide des trois documents, expliquez ce qui distingue une simple corrélation entre deux phénomènes A et B d'une relation de causalité.

On dit que deux phénomènes sont **corrélés** si, lorsque l'on observe leurs variations respectives, ils évoluent dans le même sens (**corrélation positive** : quand l'un augmente, l'autre aussi) Doc 3 électeurs d'Emmanuel Macron et niveau de diplômes - ou dans un **sens opposé** (**corrélation négative** : quand l'un augmente, l'autre diminue) – Doc 3 électeurs de Marine Le Pen et niveau de diplômes. Pour autant, cela ne signifie pas qu'il y ait un lien de causalité entre ces deux phénomènes.

La corrélation est donc un lien empirique (qui s'appuie sur une enquête, une observation et non une théorie) entre deux **phénomènes**, sans que pour autant la **variation** de l'un soit la **cause** de la variation de l'autre.

Avant de pouvoir parler d'une « bonne » corrélation, il faut constituer un échantillon pertinent et représentatif.

Si l'on souhaite établir un lien entre les pratiques électorales et le diplôme des Français, par exemple, on ne peut pas se contenter d'interroger cent personnes « au hasard » dans une rue parisienne à 12 heures. Cet échantillon ne pourra pas prétendre être représentatif de la population française : il va sans dire que les habitants et/ou travailleurs qui fréquentent une rue ou l'autre, une ville ou l'autre, à une heure ou à une autre, ne sont pas les mêmes. Les corrélations seraient donc faussées. Il est donc nécessaire de construire un **échantillon** d'observations qui doit être statistiquement **représentatif** de la population.

Il y a **causalité** lorsque la variable A est la **conséquence** de la variable B.