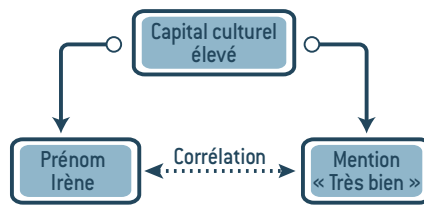


Ces questions reviennent à se demander quelle histoire causale se cache derrière la corrélation observée entre la mention et le prénom, c'est-à-dire quels sont les liens de cause à effet sous-jacents à la situation. Plus généralement, la question est : à quelles conditions des corrélations permettent-elles de tirer des conclusions causales, par exemple de conclure que le prénom Irène est la cause de la mention « Très bien » ?

Ce problème se pose quotidiennement dans tous les domaines de la science où l'on est confronté à des ensembles de données empiriques nécessitant une explication : économie, épidémiologie, toxicologie, astronomie, etc. L'interprétation correcte des corrélations observées en termes d'éventuels liens de causalité sous-jacents constitue l'un des pivots de la démarche scientifique. Elle est à ce titre d'une grande importance théorique.

Elle a aussi une grande importance pratique. C'est notamment le cas dans le domaine de la santé publique. Ainsi, quand on constate une forte corrélation entre l'exposition à une certaine substance de synthèse et une pathologie grave, il faut être sûr que cette corrélation résulte d'un lien de causalité avant que les pouvoirs publics décident d'une mesure telle que l'interdiction de la commercialisation de



2. CHEZ LES CANDIDATES au baccalauréat, le prénom Irène est corrélé à la mention « Très bien ». Mais l'existence d'un lien de cause à effet entre le prénom et la mention au baccalauréat est peu vraisemblable. La corrélation observée serait plutôt une conséquence de ce que les parents qui choisissent de prénommer leur fille Irène ont un capital culturel supérieur à la moyenne, d'où une probabilité supérieure pour leur enfant d'obtenir une bonne mention au baccalauréat.

la substance en question. Par exemple, la corrélation pourrait être due au fait que les personnes exposées à cette substance sont aussi systématiquement exposées à un autre produit, la pathologie étant causée par ce dernier et non par la substance suspectée initialement.

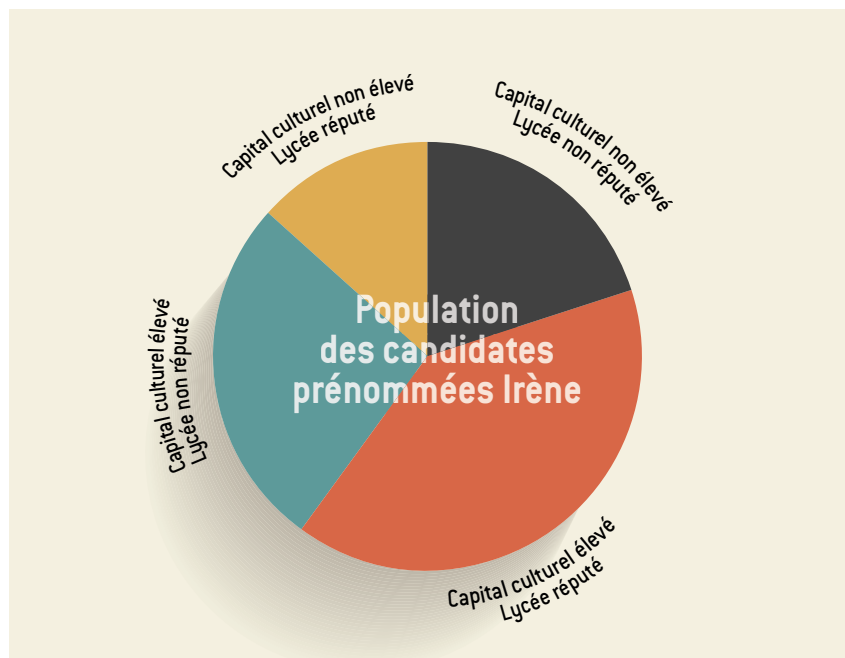
Pour passer des corrélations observées aux relations causales, deux étapes sont à franchir. La première consiste à passer des corrélations observées sur un échantillon d'une population à celles en vigueur dans la population entière. Cette première étape, la généralisation d'un échantillon de population à la population entière, relève de la science statistique. La seconde étape consiste, pour la population considérée, à passer des corrélations aux liens de cause à effet. Seule cette seconde étape concerne spécifiquement la causalité, et c'est sur elle que portera notre propos.

Diverses approches

La question de la causalité est réapparue en philosophie des sciences dans la seconde moitié du XX^e siècle. L'une des lignes d'analyse explorées visait à définir la causalité par le concept de corrélation – ou de dépendance probabiliste, selon les termes en usage chez les philosophes.

Cette façon de définir la causalité n'est pas la seule. L'une des principales définitions concurrentes repose sur l'idée que les effets n'auraient pas lieu si leurs causes n'avaient pas lieu ; cette approche peut s'appliquer à certaines études historiques par exemple. Une autre proposition, formulée dans le cadre des théorisations de la démarche expérimentale, se fonde sur la possibilité de manipuler les effets en intervenant sur les causes. Une troisième approche, enfin, est axée sur la caractérisation physique des relations causales, c'est-à-dire sur l'identification des grandeurs physiques dont la transmission de proche en proche constitue la relation de causalité.

Toutefois, l'analyse de la causalité en termes de corrélations est sans doute celle à laquelle les scientifiques font le plus appel, au moins explicitement, et c'est celle qui nous concerne ici. Cette approche repose sur l'idée qu'un effet est plus probable en présence de l'une de ses causes qu'en son absence. Ainsi, il est plus probable pour un candidat d'obtenir la mention « Très bien » s'il a un capital culturel élevé que sinon. En utilisant la notion de probabilité



3. LE PRÉNOM IRÈNE serait la cause de la mention « Très bien » si et seulement si ce prénom rendait cette mention plus probable dans chacune des sous-populations définies par les autres causes possibles de la mention. Dans l'exemple ici représenté, quatre sous-populations sont définies par le lycée fréquenté (réputé ou non) et par le capital culturel (élevé ou non).