

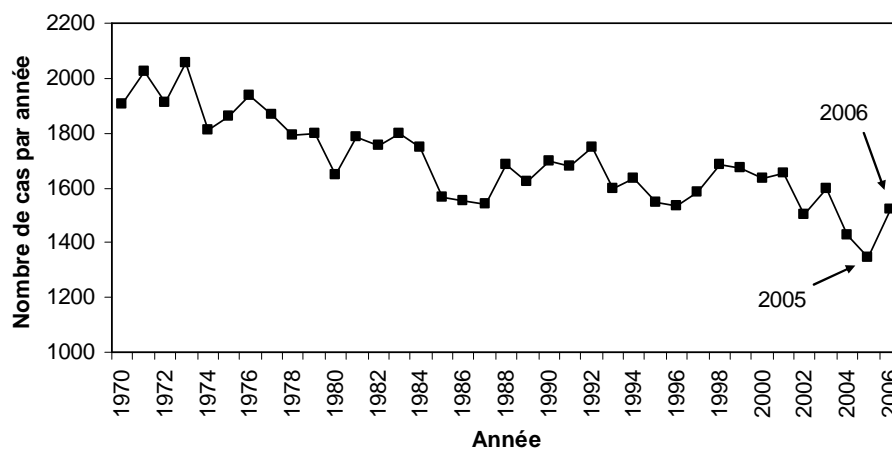
Une hirondelle ne fait pas le printemps

Stéphan Reeb
Département de biologie
Université de Moncton, Canada
© www.unregardscientifique.com

Dans les journaux, il arrive de lire des énoncés selon lesquels tel ou tel phénomène est à la hausse ou à la baisse. Souvent, ces énoncés s'accompagnent de chiffres à l'appui. Par exemple, un article pourrait annoncer que tel ou tel cancer semble devenir plus fréquent, et on verra la phrase suivante : « Ainsi, de 2005 à 2006 le nombre de cas diagnostiqués au Canada est passé de 1,345 à 1,522, une augmentation de 13% ».

L'utilisation de chiffres précis ajoute toujours du poids à un argument, mais quel est le problème potentiel avec l'exemple quantitatif ci-dessus?

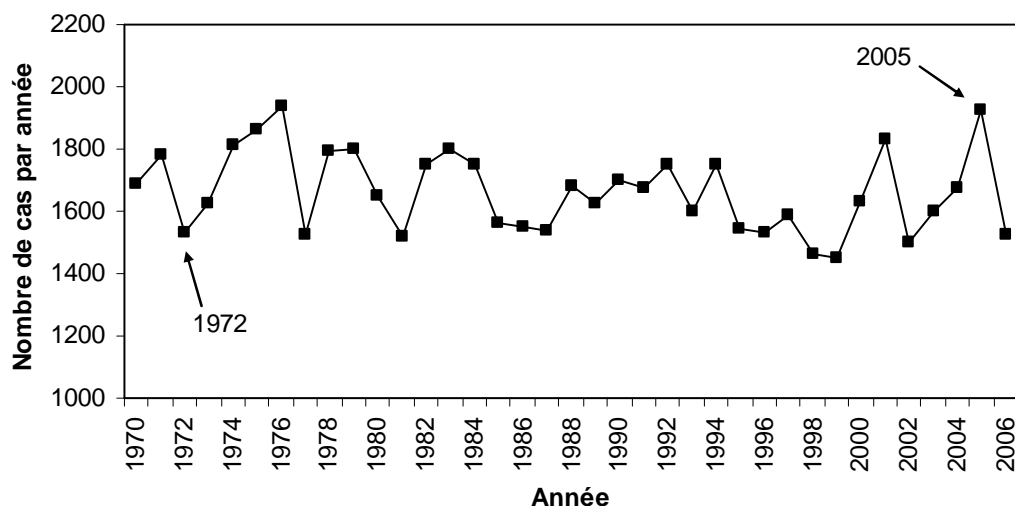
Le titre du présent document vous donne un indice. Le proverbe « Une hirondelle ne fait pas le printemps » nous exhorte à ne pas accorder trop d'importance aux événements singuliers, ou à ce qu'on appelle, dans le jargon scientifique, une faible taille d'échantillonnage. Le problème avec l'exemple ci-haut est qu'il n'implique que deux années. Or, le nombre de cas de cancer détectés annuellement, comme beaucoup de phénomènes naturels, est influencé par une multitude de facteurs, ce qui entraîne une certaine variabilité d'une année à l'autre, variabilité qui ne va pas toujours dans la même direction que la tendance générale. À titre d'exemple, jetez un coup d'œil au graphique suivant.



Comme vous pouvez le voir, il y a de la variabilité d'une année à l'autre. À la longue, le nombre de cas diminue, mais d'une année à l'autre il peut aussi bien augmenter que diminuer. Et ici la seule comparaison entre 2005 et 2006 peut nous induire en erreur, car elle va dans le sens inverse de la tendance générale. Quelqu'un qui n'aurait que les chiffres de 2005 et 2006 en main pourrait facilement tirer une conclusion générale qui en fait serait fausse.

La meilleure façon de rectifier la situation est « d'augmenter la taille d'échantillonnage », c'est-à-dire de baser notre conclusion de tendance sur un plus grand nombre d'années. Pour le scientifique, ceci revient simplement à utiliser toutes les années du graphique. Lorsque la conclusion sera publiée dans un article scientifique, il faudra que le graphique, comprenant toutes les années, soit inclus.

Pour le journaliste, la présentation d'un graphique n'est pas pratique. Bien peu de quotidiens acceptent d'incorporer des graphiques dans leurs articles. Alors que faire? Une solution est de délibérément choisir une comparaison entre deux années qui va dans la même direction que la tendance générale. Mais comment le lecteur saura-t-il que le journaliste a fait ce bon travail? Il faudra que la confiance règne. Une autre solution est d'utiliser deux années qui soient le plus espacées que possible l'une de l'autre. Mais encore là, il n'y a pas de garantie que la réalité soit bien représentée, surtout dans les cas où il n'y a pas de tendance générale, comme l'illustre le graphique suivant :



Ici, les années 1972 et 2005 sont bien espacées et elles donnent l'impression d'une augmentation, alors qu'en fait la tendance générale est nulle. (Notez qu'il faut prendre pour acquis ici que le nombre de cas est mesuré par tranche de X milliers de personnes, une standardisation qui tient compte de – qui élimine l'effet de – la hausse de population.)

D'après moi, la meilleure solution pour les journalistes est de choisir des chiffres qui impliquent le plus grand nombre d'années possible. Par exemple, on pourrait dire « Si on compare la moyenne des dix années 70s avec la moyenne des sept années de 2000 à 2006, on remarque une diminution de 18% dans le nombre de cas détectés annuellement. »

Rappelez-vous de la tyrannie des faibles tailles d'échantillonnage la prochaine fois que vous entendrez une plainte du genre « Le réchauffement de la planète mon œil! Il a fait plus froid cet hiver que l'hiver passé! »

Un problème répandu

Si un pêcheur a attrapé une truite de 15 cm dans le lac Belzébuth hier, et une truite de 20 cm dans le lac Dracula aujourd'hui, devrait-il en conclure que le lac Dracula contient de plus grosses truites? Une truite fait-elle le lac?

Si je joue quatre parties de Colons de Catane (un jeu de société que je vous recommande fortement) avec trois autres joueurs, et que je n'en gagne aucune, suis-je justifié de dire « Je suis pourri à ce jeu, car je ne gagne jamais »? Si tous les joueurs sont de force plus ou moins égale, mes chances de gagner dépendent essentiellement du hasard, et sont donc de une sur quatre. Quatre parties sont-elles suffisantes pour tester cette probabilité? Vos chances de gagner à pile-ou-face sont de une sur deux, et donc si vous perdez deux pile-ou-face en ligne, allez-vous dire « Je suis pourri à pile-ou-face; je n'y gagne jamais »?

Si quelqu'un de Moncton mentionne à un ami de Montréal que son voisin est alcoolique et qu'il est sans emploi, et que l'ami répond que – quelle coïncidence! – lui aussi il a un voisin qui est alcoolique et sans emploi, seront-ils tentés de conclure que l'alcoolisme mène au chômage? Même si on oublie la difficulté d'établir une relation cause-à-effet ici – peut-être que c'est le chômage qui est arrivé en premier et qui a mené à la déprime et l'alcoolisme, ou peut-être que l'alcoolisme et le chômage ne s'influencent pas l'un l'autre mais sont tous les deux la conséquence d'un troisième facteur, comme la pauvreté – ne devrait-on pas s'inquiéter du fait que cette coïncidence n'implique que deux personnes, avant de tirer des conclusions?

Si l'employé d'un zoo fait l'hypothèse que les chimpanzés aimeraient mieux manger des patates cuites plutôt que des patates crues, et qu'il fait une petite expérience où il offre un choix aux trois chimpanzés de son zoo entre un plat de patate cuites et un plat de patates crues, et qu'il observe que chacun des trois chimpanzés mangent plus de patates cuites que crues, est-il alors convaincu que son hypothèse est vraie? Avant d'appliquer une conclusion à une espèce entière, cet employé à l'esprit scientifique ne

devrait-il pas s'inquiéter du fait que seulement trois chimpanzés ont été testés, et seulement une fois chacun?

Du travail supplémentaire pour les scientifiques

Les questions ci-haut reflètent toutes une interrogation centrale : combien d'hirondelles est-ce que ça prend pour faire le printemps? Combien d'échantillons doit-on prendre pour se convaincre que telle relation ou telle différence est réelle et non pas causée par le simple hasard ou par la variabilité induite par d'autres facteurs qui ne nous intéressent pas pour l'instant?

Il existe une branche des mathématiques – la statistique – qui s'occupe de répondre à cette interrogation. Grâce aux recherches des statisticiens, il est habituellement possible de déterminer, avec un degré de confiance quantifiable, et connaissant la taille d'échantillonnage, si telle relation ou telle différence entre deux variables est bel et bien réelle ou simplement dûe au hasard. Il n'est donc pas surprenant de constater que la formation scientifique comprend souvent des cours en statistique. Par exemple, au Département de biologie de l'Université de Moncton, il existe au premier cycle un cours de « Biométrie » et au deuxième cycle un cours de « Dispositifs expérimentaux », tous deux portant sur les tests statistiques couramment employés en sciences de la vie.

(Je me dois ici de mentionner que les sciences physiques et le génie dépendent moins des statistiques, puisque ces disciplines œuvrent souvent avec des systèmes qui sont moins complexes, avec moins de facteurs causaux, que les sciences de la vie.)

En plus de rajouter à la liste de cours que doivent prendre les étudiants, la problématique de l'échantillonnage suffisant donne du travail aux scientifiques. Dans les reportages de science vulgarisée à la radio ou à télévision, on entend souvent le chef de laboratoire annoncer la conclusion de son étude et décrire les grandes lignes de son approche expérimentale. Mais derrière ces trois minutes médiatiques, il se cache des dizaines, peut-être même des centaines d'heures passées par des étudiants de maîtrise ou de doctorat, ou par des employés contractuels, à mesurer un très grand nombre d'échantillons. Faire de la science engendre souvent de brefs moments très intenses et gratifiants – le graphique qui apparaît sur l'écran confirmant notre hypothèse originale; la réalisation soudaine de l'effet qui explique tous nos résultats – mais soyons honnête : la science comporte aussi souvent un travail de routine long et pénible visant simplement à obtenir la confirmation statistique de ce qui nous était apparu évident (bien sûr!) dès le premier échantillon.

Mais l'hirondelle peut être la première d'une vague!

La morale du présent discours est qu'il ne faut pas construire des conclusions à partir d'un nombre limité de cas. Cependant, n'allez pas croire que les événements rares sont pour autant sans valeur. Premièrement, de tels événements peuvent servir de base à l'élaboration d'hypothèses de travail. Par exemple, quelques heures avant que l'ouragan Gabrielle ait frappé la Floride en septembre 2001, les requins bordés (« blacktip shark ») qui vivaient près de la côte se sont déplacés vers le large, comme s'ils anticipaient et essayaient d'éviter les grandes vagues de l'ouragan. Une équipe de chercheurs dirigée par l'ichtyologue Michelle Heupel a trouvé que le moment précis du départ des requins coïncidait avec la baisse de pression atmosphérique qui a précédé l'arrivée de Gabrielle. Cette observation unique a amené les chercheurs à émettre l'hypothèse – attention je dis bien l'hypothèse, pas la conclusion – que les requins pouvaient ressentir les baisses de pression atmosphérique même sous l'eau.¹ Cette hypothèse a par la suite fait l'objet de travaux expérimentaux plus poussés en laboratoire, et d'observations supplémentaires lorsque l'ouragan Charley a frappé la même région en 2004, et les résultats ont jusqu'à maintenant supporté l'hypothèse.

Deuxièmement, certains événements singuliers peuvent s'avérer concordants avec d'autres types d'évidence et ainsi ajouter leur modeste support à une hypothèse déjà existante. Par exemple, la panne d'électricité qui s'est étendue sur l'ensemble du nord-est américain et de l'est du Canada pendant environ deux jours en août 2003 a empêché la plupart des usines de charbon de fonctionner. Des chercheurs de la University of Maryland ont profité de cette situation unique pour prendre des échantillons d'air et comparer ces échantillons avec d'autres pris la même journée à des endroits non touchés par la panne de courant, ou avec des échantillons pris précédemment au même endroit lorsque les usines étaient en opération. Ils ont mesuré des baisses de 90% et 50 % dans les niveaux de dioxyde de soufre et d'ozone respectivement, confirmant ainsi directement la grande part que joue les usines de charbon dans la production de polluants, laquelle avait déjà été supposée à partir de mesures d'émission et de modèles numériques. Comme la circulation automobile n'avait pas tellement changé lors de la panne de courant, les chercheurs ont aussi avancé l'hypothèse que la contribution polluante des usines de charbon par rapport à celle des voitures était plus grande qu'on ne l'avait pensé auparavant.²

Les faibles tailles d'échantillonnage et les événements singuliers peuvent donc trouver leur place en science. L'important est de ne pas tirer de conclusions seulement à partir d'eux.

¹ Heupel, M.R., C.A. Simpfendorfer and R.E. Hueter (2003). Running before the storm: blacktip sharks respond to falling barometric pressure associated with Tropical Storm Gabrielle, *Journal of Fish Biology* 63: 1357-1363.

² Marufu, L. T., B. F. Taubman, B. Bloomer, C. A. Piety, B. G. Doddridge, J. W. Stehr, et R. R. Dickerson (2004). The 2003 North American electrical blackout: An accidental experiment in atmospheric chemistry, *Geophysical Research Letters* 31: L13106, doi:10.1029/2004GL019771.