

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

DANIEL SCHWARTZ

Statistique et vérité

Journal de la société statistique de Paris, tome 125, n° 2 (1984), p. 74-83

http://www.numdam.org/item?id=JSFS_1984__125_2_74_0

© Société de statistique de Paris, 1984, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

STATISTIQUE ET VÉRITÉ (*)

Daniel SCHWARTZ

Professeur, directeur de l'Unité de recherches statistiques de l'Institut national de la santé et de la recherche médicale, ancien président des Sociétés de statistique

Dans le domaine de l'aléatoire, où le risque d'erreur est, de principe, inévitable, la démarche statistique est un puissant auxiliaire de « la vérité », en permettant une conclusion objective, avec un risque d'erreur contrôlé. Comment se fait-il alors qu'elle soit si souvent considérée comme mensongère? La raison majeure est un dérapage, principalement à trois étapes du raisonnement : le jugement de signification, l'interprétation, en particulier sous forme d'une relation causale, et l'énoncé des conclusions. De tels dérapages sont illustrés par de nombreux exemples.

In the area of uncertainty, where the risk of error is, theoretically, inevitable, statistical procedures are a powerful auxiliary tool in determining the « truth », in that they make possible an objective conclusion with a controlled risk of error. Why then are statistics so often considered as deceitful? The chief reason is a drift or a foreseeable modification that may occur in three major steps of reasoning: (1) evaluation of significance, (2) interpretation, especially in terms of causal association, and (3) statement of conclusions. Numerous examples illustrate such drifts.

Ainsi donc la proposition me fut faite, un beau jour, d'inclure la Statistique dans un volume consacré à la Vérité, le titre en serait « Statistique et Vérité ». J'en eus le souffle coupé. Était-ce une provocation? Statistique et Vérité : l'adultère absolu. La statistique est, chacun le sait, indéfectiblement mariée au mensonge.

Quelle tentation! Plaisir d'associer, ne fût-ce que le temps d'un article, le mot Statistique au mot Vérité. Plaisir qui devient vite incertain, hésitant, car s'il y a matière à article, c'est que cette liaison doit poser des problèmes? Eh bien que ce soit alors l'occasion, pour le statisticien, de les aborder franchement, de les décortiquer, d'en expliquer le pourquoi, et par quels malentendus il fait inexorablement figure de menteur.

Je laisserai de côté la Statistique descriptive — les statistiques — et les innombrables sources d'erreur provenant du seul fait de données inexactes ou imprécises. Le lecteur pourra trouver sur ce sujet un très bel exposé de J. Klatzmann ⁽¹⁾. Et j'axerai ma réflexion sur le raisonnement statistique dans la recherche, en choisissant comme domaine la recherche médicale.

L'impossible certitude

Un exemple type pour illustrer un constat d'impuissance. Dans une race de souris, la fréquence des cancers spontanés est bien connue, mettons 40 %. C'est dire que, pour chaque souris, le résultat dépend d'une loterie avec 40 % de billets « cancer »; sur un échantillon de 100 souris le nombre de can-

(*) Communication faite le 21 décembre 1983 devant les Sociétés de statistique de Paris et de France. Le texte de cette communication a été publié dans « *Le genre humain* », 7-8 « La Vérité », Éditions Complexe, Bruxelles, 1983.

1. J. Klatzmann, « Les pièges de la statistique », dans les *Cahiers Rationalistes* 188, juin juillet 1960, 24, rue des Grands-Augustins, PARIS-6^e.

céreuses varie, d'une fois à l'autre, autour de 40, selon les « fluctuations d'échantillonnage », c'est-à-dire les caprices du hasard. De vrais caprices, imprévisibles et parfois grandioses : 0 ou 100 cancers sont des résultats, certes improbables, mais pas impossibles.

Voici maintenant qu'un biologiste veut savoir si sa dernière invention, la drogue D, modifie cette fréquence (dans un sens ou dans l'autre). A cet effet il entreprend une expérience, administrant la drogue à, par exemple, 100 souris. Il observe un certain pourcentage de cancers. Comment va-t-il conclure?

La démarche adoptée pendant des siècles était la suivante : la drogue sera dite active ou inactive selon que l'écart entre la fréquence observée et 40 % est grand ou petit. Mais que veut dire « grand »? Que veut dire « petit »?

Si la drogue est inactive, le pourcentage observé fluctue autour de 40 %, mais les caprices du hasard pourront conduire très loin de 40 %. Inversement, si la drogue est active et fait passer la probabilité de cancer à, par exemple, 30 %, la fréquence observée sur 100 souris va fluctuer autour de 30 %, mais les caprices du hasard pourront conduire tout près de 40 %. Ainsi drogue active ou inactive ne veut pas dire *forcément* écart grand ou petit. La conclusion de cette réflexion est très claire : il est impossible de se prononcer *avec certitude* : la « vérité » est hors de portée. Et la raison de cette impossibilité est également claire : c'est que l'apparition du cancer est un événement *aléatoire*, régi par une loterie faisant varier capricieusement les résultats d'une série à l'autre. Si l'événement était *certain* — toutes les souris sont cancéreuses, ou au contraire indemnes — nous n'aurions pas cette difficulté.

Qui dit aléatoire dit variabilité : certaines souris ont le cancer, d'autres ne l'ont pas. Or la variabilité affecte tous les aspects du vivant : taille, taux de cholestérol, réaction au tabagisme, durée d'incubation d'une maladie, survie d'un cancéreux, issue d'un traitement... Dès lors la difficulté rencontrée pour les souris va surgir à chaque pas : qu'on veuille comparer la fréquence des cancers chez les fumeurs et les non-fumeurs (problème étiologique), la durée de survie de cancéreux selon que les ganglions sont ou non envahis (problème de pronostic), les pourcentages de guéris avec deux traitements (problème de thérapeutique), les taux ou les moyennes à comparer sont soumis aux fluctuations d'échantillonnage, on ne peut donc au vu de deux séries conclure avec certitude. Constat formel, déconcertant : *dans la quasi-totalité des problèmes posés par la recherche médicale, une réponse certaine est, de principe, exclue*. Constat qui s'étend à toutes les sciences de la vie et, d'une manière générale, au domaine de l'aléatoire.

C'est dans cette situation d'impossible certitude qu'est intervenue la statistique.

La démarche statistique

Le hasard a ses caprices, mais aussi ses lois : les lois de probabilité. Sur nos 100 souris, la fréquence des cancers spontanés peut théoriquement s'écarter beaucoup de 40 %, mais un grand écart est improbable, d'autant plus improbable qu'il est plus grand. Le tout est de chiffrer cette relation. Le calcul des probabilités indique que l'écart a, par exemple :

5 chances sur 100 de dépasser 10 %,

1 chance sur 1 000 de dépasser 17 %.

Administrons maintenant la drogue D à 100 souris. Supposons la drogue inactive, seuls apparaîtront des cancers spontanés, et les écarts à 40 % seront régis par les règles illustrées ci-dessus. Décidons alors de ne considérer un écart comme *significatif* — la drogue D comme active — que si l'écart à 40 % dépasse 10 %, comme non significatif s'il est plus faible, c'est-à-dire qu'on ne conclura pas si le pourcentage tombe entre 30 et 50 %. Cette politique conduira certes à des erreurs : avec un produit inactif, l'écart dépassera quand même, de temps à autre, 10 %, et on le déclarera actif; mais ce risque d'erreur est connu, c'est 5 %. Comme il est chagrinant de se tromper, on peut adopter une politique plus sévère, et déclarer l'écart significatif s'il dépasse, par exemple, 17 % : le risque d'erreur passe à 1 pour 1 000. Mais l'intervalle de non-conclusion passe à 40 % \pm 17 %, beaucoup plus grand que le précédent, on a

beaucoup plus de chances de laisser échapper un produit actif : à vouloir diminuer un risque, le risque d'erreur, on en augmente un autre, le manque à gagner, c'est le dilemme classique de tout jugement basé sur des probabilités, ainsi en justice s'affrontent les risques d'erreur judiciaire et d'acquiescement du coupable. On ne peut donc diminuer à volonté le risque d'erreur, il faut s'arrêter à une valeur raisonnable. Mettons que ce soit, comme on l'a supposé en premier, 5 %. Le statisticien qui adoptera cette politique verra limité à 5 % le pourcentage des problèmes où il énoncera une contre-vérité.

Un point très éclairant est la relation entre l'écart significatif et l'effectif; dans l'exemple des souris, avec le risque 5 %, le calcul montre qu'un écart sera significatif s'il dépasse :

— pour un échantillon de 100 souris... 10 % (indiqué ci-dessus)

— pour un échantillon de 1 000 souris... 3 %.

Ainsi, avec 37 % de cancers (écart 3 %) la drogue sera déclarée active si l'expérience comportait 1 000 souris, mais pas si elle en comportait 100 : un écart de 3 % est donc « grand » sur 1 000 souris, « petit » sur 100. C'est dire combien était subjective, et mauvaise, la démarche pré-statistique, où on concluait à l'activité ou à la non-activité selon que l'écart était grand ou petit, notion qui n'a aucun sens, et quelle est la supériorité de la méthode statistique qui lui substitue la notion objective d'écart significatif.

La notion d'écart significatif peut être davantage encore quantifiée, en exprimant si un écart est plus ou moins significatif. Si, avec 100 souris, on observe 23 % de cancers (écart 17 %) le calcul montre, comme indiqué ci-dessus, qu'on a moins de 1 chance sur 1 000 d'observer un écart de cet ordre, l'écart est dit « significatif à 1 pour 1 000 ».

Statistique et vérité

Dans un domaine où le risque d'erreur est, de principe, inévitable, voilà donc le progrès, la révolution : sans la démarche statistique, l'erreur résulte d'une décision subjective avec une fréquence incontrôlée; avec la démarche statistique, la décision est objective, son risque d'erreur plafonné, son degré de signification connu. La démarche statistique est un puissant auxiliaire de la vérité. Entre la statistique et la vérité, c'est le premier point de rencontre, un point fort totalement à l'actif de la statistique, un article sur Statistique et Vérité pourrait s'arrêter là. Mais l'autre volet s'impose, qui pourrait s'appeler Statistique et Mensonge : pourquoi cette indéfectible association?

Statistique et mensonge

La première raison est tout simplement le revers de la médaille. D'avoir pactisé avec le mensonge, la statistique sort compromise, comme un agent double. Réglementer l'erreur, c'est l'admettre. Bien plus, la statistique la cautionne : la démarche se veut scientifique, elle utilise un test bien élaboré, avec des formules mathématiques — mathématiques donc sûres, n'est-ce pas? — et le résultat est quand même, de temps à autre, une contre-vérité. Voilà bien un passage du mensonge à une forme raffinée du mensonge. Ce motif d'accusation est singulier; il doit être mis à part, c'est une réaction purement psychologique, qui fait endosser à la Statistique le péché originel inhérent au domaine où on l'utilise, mais c'est injuste, la statistique n'est pas fautive, il n'y a aucune erreur de raisonnement. Tandis que les autres raisons d'accuser la statistique concernent au contraire des dérapages dans le raisonnement.

De tels dérapages sont innombrables et variés, mais pour ne retenir que les principaux et dans un premier classement, on peut considérer trois niveaux : un phénomène est-il significatif? Si oui, comment l'interpréter, notamment, sur le plan de la causalité? Comment l'énoncer et l'utiliser? Seul le premier niveau est spécifique de la démarche statistique proprement dite, au sens où nous venons de la définir, les deux autres s'étendent à l'utilisation « des statistiques » voire à des aspects généraux du raisonnement déductif. Le premier niveau sera étudié après les deux autres. Pourquoi cette entorse à la logique

et à la chronologie? Dans ce numéro, dans cet article, quand on est statisticien, mieux vaut avouer la vérité : cette partie est plus difficile à expliquer et à comprendre, on eût risqué de rebuter le lecteur à peine engagé...

I. — INTERPRÉTATION DES RÉSULTATS

Le jugement de causalité

Si les résultats sont significatifs — et dans ce cas seulement — reste à les interpréter. La plus fréquente erreur est, de beaucoup, le passage injustifié à l'interprétation causale. Mais la notion de cause demande d'abord définition.

Une généralisation de la notion de cause

Il est admis que le tabac est cause du cancer bronchique. Mais la plupart des fumeurs y échappent, et des non-fumeurs en sont victimes. Voilà qui heurte deux idées profondément ancrées : celles de l'effet certain et de la cause unique. D'autant plus ancrées, dans le domaine médical, qu'elles ont été longtemps confortées par le triomphe de l'idéologie Pastoriennne : une maladie = un microbe (à lire dans les deux sens). A s'en tenir là, on sait bien maintenant qu'on ne pourrait pas attribuer de cause à la majorité des maladies, y compris les maladies infectieuses. En vertu de la variabilité qui caractérise fondamentalement le vivant, les sujets exposés à un facteur de risque auront quelquefois la maladie, mais pas toujours. Et en vertu d'un déterminisme plurifactoriel des maladies, les sujets non exposés à un facteur pourront la contracter du fait d'un autre. De sorte qu'il a fallu généraliser la notion de facteur causal d'un événement : il suffit qu'il cause une augmentation de sa probabilité de survenue. C'est, en quelque sorte, une « relation de cause à probabilité d'effet ». Mais que le facteur soit bien la cause de cette augmentation de probabilité n'est pas si facile à prouver, c'est là qu'est la source de bien des dérapages.

Tabac et cancer

De très nombreuses enquêtes l'ont montré, confirmé à satiété : la fréquence du cancer bronchique est très significativement et très nettement plus élevée chez les fumeurs. Mais voilà, les fumeurs diffèrent-ils des non-fumeurs par le seul fait qu'ils portent une cigarette à la bouche? Ce serait bien curieux. Les sujets se sont choisis eux-mêmes fumeurs ou non-fumeurs. On peut donc suspecter entre eux toutes sortes de différences, que des enquêtes ont confirmées : comparés aux non-fumeurs, en moyenne bien sûr, les fumeurs diffèrent par des traits sociaux et psychologiques, la situation familiale, la consommation d'alcool et de café, ils mesurent même... 1 cm de plus. Alors, la cause du cancer, est-ce le tabac ou un de ces nombreux facteurs?

Les déboires d'un jeune accoucheur

Il avait décidé, pour sa thèse de médecine, de comparer le taux de complications à la naissance selon que la femme fait ou non appel à l'accoucheur. Le résultat le surprit : les complications étaient plus fréquentes quand l'accoucheur était présent. Alors seulement il réfléchit, et comprit : les femmes appelaient davantage l'accoucheur dans les cas graves. De même que celles des fumeurs et des non-fumeurs, les deux séries n'étaient pas comparables; derrière le facteur « présence de l'accoucheur » un autre était dissimulé, la gravité du cas. C'est comme pour les trains dans les passages à niveau : un facteur peut en cacher un autre...

Soixante-dix pour cent des gens meurent au lit...

Donc ne vous couchez pas. Dans cet aphorisme célèbre, l'erreur est la même que dans les deux exemples précédents. Si je compare, aujourd'hui à midi, les gens couchés et debout, les deux séries ne diffèrent pas seulement par la position horizontale ou verticale, la première comporte davantage de malades. L'erreur est ici poussée à l'extrême, la cause et la conséquence sont inversées.

Des exemples par dizaines

Des exemples analogues aux précédents, on pourrait en citer par dizaines. On en présentera seulement quelques-uns, illustrant leur diversité.

— Certaines enquêtes (mais pas toutes) ont fait apparaître une fréquence du cancer du sein plus élevée chez les femmes qui prennent la pilule. Mais celles-ci se distinguent des autres sous de nombreux aspects, notamment leur vie sexuelle et reproductive, qui intervient précisément dans l'étiologie de ce cancer.

— Les femmes irradiées pendant leur grossesse ont un peu plus souvent des enfants leucémiques. Mais si elles ont été irradiées, c'est parce qu'elles présentaient telle ou telle pathologie. Quelle est la cause de la leucémie, l'irradiation, la pathologie qui l'a nécessitée, un des facteurs qui ont conduit à cette pathologie?

— Une histoire lue dans un vieux numéro d'une revue éditée par des étudiants en psychologie. On y compare le taux de succès à l'examen de statistique, selon que les élèves ont ou non assisté aux séances de travaux dirigés. Minutieusement, les étudiants font le test statistique, la différence est significative, et dans le sens conforme au bon ordre des choses. Et de conclure, à l'adresse des conscrits : suivez les travaux dirigés! Qui dira si les assidus n'étaient pas, en tout état de cause, les « meilleurs »?

— Enfin un exemple emprunté au domaine de la thérapeutique. Les angiomes tubéreux cutanés des enfants sont usuellement traités par irradiation. Mais comme la plupart disparaissent spontanément, l'utilité du traitement fut, il y a quelque vingt ans, remise en question. Un radiologue organisa une réunion, où il projeta des images d'angiomes avant et après irradiation. L'« après » était toujours spectaculairement meilleur que l'« avant ». Qui pouvait résister à pareille démonstration? Le statisticien, bien sûr. Les séries « avant » et « après » ne différaient pas seulement par le traitement, elles différaient aussi par un facteur capital, le temps, qui précisément pouvait avoir fait son œuvre tout seul!

Les limites de l'enquête d'observation

Le point commun à tous les exemples précédents est qu'on n'a pas déterminé les groupes à comparer, ils se sont déterminés d'eux-mêmes : ce ne sont pas des expérimentations, ce sont des enquêtes d'observation. Alors le facteur qui détermine les groupes, — fumer, appeler l'accoucheur, être au lit... — est vraisemblablement lié à beaucoup d'autres, les groupes diffèrent pour toute une série de facteurs, on ne sait plus lequel incriminer. En bref, *la tare foncière de l'enquête d'observation est l'impossibilité de principe de prouver une relation de causalité. Et c'est l'ignorance de cette règle élémentaire qui constitue une des principales sources d'erreur — de « mensonge » — en statistique.*

Mais alors, que faire?

Voilà qu'un autre danger menace le statisticien : l'inefficacité, pour éviter le mensonge; car enfin va-t-il choisir qui fume, quelle femme sera irradiée pendant la grossesse? C'est poser le problème des possibilités et des limites de la démarche statistique. On peut assez facilement les résumer.

Quand l'expérimentation est possible, il faut absolument y recourir. C'est le cas en thérapeuti-

que, où la démarche statistique a apporté encore un progrès considérable, en prescrivant qu'il fallait comparer les traitements sur des séries déterminées par tirage au sort, quand l'éthique le permet. Les groupes sont alors aussi comparables que possible, et s'il y a différence dans les résultats, elle est bien due aux traitements.

Quand l'expérimentation n'est pas possible, la preuve de la causalité ne peut être administrée par la seule enquête d'observation mais, dans le cas du tabac par exemple, on a pu, par une longue accumulation d'arguments cohérents, incluant des expériences sur animal, arriver à une quasi-certitude : c'est une démarche coûteuse, mais possible quand l'enjeu en vaut la peine.

Dans d'autres cas, on peut se fixer un objectif moins ambitieux et plus rentable : une aide à la recherche causale. Ainsi, dans l'étiologie du mongolisme, on a observé que la fréquence augmente avec l'âge de la mère; mais les mères âgées diffèrent des plus jeunes par au moins deux facteurs : l'âge plus élevé du père, le rythme plus faible des rapports sexuels. Ce dernier facteur pourrait être important, car il entraîne une attente des gamètes au rendez-vous de la fécondation, attente qui chez l'animal élève le taux de malformations. La méthode épidémiologique a permis de montrer que l'âge de la mère intervient en propre, indépendamment de ces deux autres facteurs; il n'est pas pour autant « la cause », et la recherche doit être poursuivie, mais la situation a été nettement clarifiée.

Un pas de plus dans la réduction des exigences est celui du renard et des raisins : le renoncement. Il n'est pas synonyme d'inefficacité, car il est heureusement possible, dans de multiples domaines, d'agir sur un événement sans en connaître les causes. Ainsi, en périnatologie, on peut, en combinant de multiples facteurs, identifier des grossesses à haut risque; on ne se préoccupe pas de savoir si ces facteurs ont un rôle causal, beaucoup d'entre eux sont d'ailleurs non modifiables — l'âge de la femme par exemple! — mais on peut prendre une mesure totalement indépendante de ces facteurs : surveillance régulière, accouchement dans une maternité bien équipée. Qu'on ne s'étonne pas si une telle « *épidémiologie pragmatique* » se développe chaque jour davantage : elle est efficace et met à l'abri d'une source majeure de dérapage dans le raisonnement, puisqu'elle évite le jugement de causalité.

II. — L'ÉNONCÉ DES CONCLUSIONS

Vérité aujourd'hui, mensonge demain

C'est une anecdote imaginaire (2). Un jour deux pays A et B comparèrent leur mortalité annuelle. Elle était plus faible dans le pays B, qui s'en réjouit à grand fracas de panneaux publicitaires : « on meurt moins chez nous ». — « Erreur, objecta le gouvernement de A, la pyramide des âges n'est pas la même dans nos deux pays, à âge égal notre mortalité est plus faible. » Les panneaux publicitaires « on meurt moins chez nous » changèrent de camp. Le pays B appointa un statisticien, qui réfléchit et observa que le climat était plus sévère dans son pays; à égalité d'âge et de conditions climatiques, B était gagnant, on y remit les panneaux. Mais le pays A créa un Institut de Statistique, doté de moyens substantiels, et l'Institut remarqua que la densité médicale était plus faible dans A. Qui ne sait l'importance de ce facteur, et qu'en l'absence de médecin nous mourons comme des mouches? Il était juste, nécessaire, d'en tenir compte. A égalité d'âge, de conditions climatiques et de densité médicale, la mortalité s'avéra... disons, parce qu'il faut bien finir l'histoire, la même dans les deux pays. Et ceux-ci, désireux d'arrêter là les frais, organisèrent une grande fête, avec un bûcher où on jeta les statisticiens des deux pays. Les statisticiens, ces menteurs, qui avaient fait dire, au fil du temps, une chose et son contraire.

2. L'idée en est empruntée à une conférence de Gini.

En fait la contradiction provenait, non des calculs statistiques, mais de l'énoncé des conclusions. La mortalité plus faible dans B était un résultat indiscutable, et qui ne pouvait être remis en question. Les résultats suivants devaient être énoncés sous la forme : la mortalité de tel pays est inférieure à égalité de tels facteurs. Supprimer cette précision est un mensonge par omission. Mensonge d'autant plus troublant qu'il correspond à des progrès dans la connaissance (on a mis en évidence des facteurs de mortalité), et des pas successifs vers une certaine vérité, la démarche effectuée tendant à comparer les deux pays dans des conditions aussi comparables que possible. En fait la comparaison de deux groupes, qui est l'essentiel de la recherche en statistique médicale, peut presque toujours être envisagée sous deux aspects : globale, ou à égalité du plus grand nombre de facteurs possible; la première démarche est plutôt pragmatique, la deuxième explicative. Elles doivent différer par l'énoncé des conclusions, mais aussi, bien avant, par la formulation du problème.

Une histoire de haricots

Je revois encore un jeune agronome me montrant, sur le terrain, un plan expérimental sophistiqué entremêlant des parcelles de haricots de races A et B, dont on voulait comparer le rendement. Les plantes A étaient tout en hauteur, les plantes B, très basses, s'étalaient sur le sol et se gênaient mutuellement. Je l'interrogeai sur la densité de plantation adoptée. « La même pour A et B, bien sûr. » — « Pourquoi? » — « Pour les comparer à conditions égales. » Que venait faire ici l'égalité? Le problème n'a aucune visée explicative, il est purement pragmatique, il s'agit de choisir la race qui donne le meilleur rendement dans des conditions optimales — pour elle — de densité, d'engrais, etc. Cet exemple illustre non seulement l'ambiguïté de l'énoncé « A meilleur que B » quand on ne précise pas les facteurs pris en compte, mais aussi la nécessité de détruire cette ambiguïté par une formulation correcte au départ.

La comparaison de deux traitements

On a parlé plus haut des angiomes tubéreux cutanés, usuellement traités par irradiation, mais qui souvent disparaissent seuls. L'irradiation est-elle efficace? Pour le savoir, on constitua, en 1960, deux groupes, traité et témoin, par tirage au sort. C'était en France le premier essai thérapeutique utilisant cette méthode. L'irradiation s'avéra efficace. Cependant on constata plus tard que le résultat était vrai pour les garçons, mais pas pour les filles. Encore une vérité variable au cours du temps!

Cet exemple peut paraître une illustration bien banale d'un simple progrès de la connaissance. Il recouvre en fait un problème très général en statistique : la conclusion d'un essai thérapeutique, et de bien d'autres problèmes, devrait en toute rigueur être énoncée : A est *en moyenne* meilleure que B. Ce qui bien souvent heurte le médecin : « Qu'en ai-je à faire, de votre vérité moyenne, qui ne serait pas la vérité pour chacun? » La réponse est claire. Si on suppose que la conclusion peut être inversée selon certains facteurs, sexe, âge, forme clinique (ce n'est pas si fréquent), il faut les inclure dans l'essai, on peut alors déceler des conclusions diversifiées selon les catégories de malades. Si on ne l'a pas fait, alors le traitement A, parce qu'il est meilleur en moyenne pour l'ensemble des malades, doit être administré à chacun.

Faut-il manger maigre?

Une enquête épidémiologique célèbre montra, vers le milieu de ce siècle, que le risque cardiovasculaire augmentait fortement avec le taux de cholestérol. On recommanda d'éviter les graisses. Des études de plus en plus sophistiquées incriminèrent ensuite certaines graisses seulement, puis les proportions relatives de diverses graisses, et des consignes alimentaires furent successivement diffusées, par les

médecins et les médias, consignes de plus en plus complexes et contredisant les précédentes. Le crédit des statistiques commençait à s'en ressentir fortement, quand un inquiétant résultat surgit d'une enquête récente : plus la cholestérolémie était basse, plus fort était le risque de cancer. Belle récompense pour ceux qui s'étaient, durant des années, privés de beurre, de fromage et de crème!

Cette histoire illustre la succession de vérités, ou de mensonges, classique dans le domaine de l'alimentation, et le profond scepticisme qui en résulte. Elle s'explique par plusieurs causes.

La première est le progrès des connaissances, dont on ne saurait se plaindre. Peut-être seulement soulignera-t-on qu'un renversement des idées est généralement considéré comme une découverte, sauf en statistique où c'est une preuve de mensonge...

La seconde est un énoncé trop simpliste des conclusions : ériger en règle « il faut manger ceci ou cela » suppose un bilan coût-avantages en l'état actuel des connaissances, qui en général n'est pas fait.

Enfin, comme rien n'interdit le cumul de plusieurs dérapages, on ajoutera que, résultant d'enquêtes d'observation, la conclusion causale était infondée; si elle est plausible, pour certaines graisses, dans le domaine cardio-vasculaire, elle était par contre franchement fausse pour le cancer : on s'aperçut facilement que la cholestérolémie basse était l'indice de la cachexie à venir, c'était la conséquence d'un cancer et non sa cause, voir « 70 % des gens meurent au lit »...

En conclusion

Un mauvais énoncé des conclusions est un point de dérapage vers le mensonge particulièrement fréquent, subtil et dangereux. Il peut résulter d'une simple omission. Polymorphe, il est difficile à catégoriser. Mais, pour l'essentiel, il résulte d'une extrapolation injustifiée de résultats obtenus dans des conditions précises, qu'il convient d'énoncer. Parmi celles-ci figurent en tout premier lieu les facteurs pris en compte, liés à une recherche d'égalisation, qui peut être souhaitable ou non selon la formulation du problème. Le fait qu'on raisonne sur des moyennes est également important, il n'est d'ailleurs qu'une façon d'ignorer les tiers facteurs.

III. — LE JUGEMENT DE SIGNIFICATION

Le jugement de signification contrôle le risque d'erreur, mais à condition de respecter une règle du jeu, qui est assez subtile.

Quel est le risque d'erreur?

Un statisticien décide, pour l'écart significatif, d'adopter le risque 5 %. Cela veut-il dire qu'il se trompera, en moyenne, 5 fois sur 100? Non. Voici en effet, en reprenant l'exemple des souris à cancer, le nombre d'erreurs attendu, sur une série de drogues présentées, en comportant n inactives et n' actives :

- pour les n drogues inactives : $n \times 5 \%$,
- pour les n' drogues actives : 0 (en effet, ou le statisticien conclura « actif » et il aura raison, ou il ne conclura pas, *ce n'est pas une erreur, mais un « manque à gagner »*),
- soit, au total, pour les $(n + n')$ drogues : $\frac{n}{n + n'} \times 5 \%$.

Si le statisticien travaille avec un biologiste qui a du flair, et ne lui apporte que des drogues actives ($n = 0$), le risque d'erreur sera nul. Si le biologiste est un piteux chercheur, et n'apporte jamais de produits actifs ($n' = 0$), le risque sera de 5 %. La réalité sera bien sûr entre ces deux extrêmes, mais il s'agit

le plus souvent de soumettre à expérience des drogues plausiblement actives, le rapport $n/(n + n')$ sera faible, le risque d'erreur inférieur à 5 %. Transposé de cet examen de drogues à un ensemble de phénomènes soumis à vérification statistique, phénomènes traduisant le plus souvent des hypothèses plausibles, le risque d'erreur est nettement inférieur à 5 %. Ce qui est fort heureux, car admettre 5 % d'erreurs, vu l'énorme quantité de recherches effectuées, conduirait à un nombre de publications erronées tout à fait inacceptable. Mais cette conclusion suppose qu'on évite des dérapages.

Premier dérapage : l'absence d'hypothèse

Les ordinateurs conduisent les chercheurs à introduire un nombre considérable de variables, et à croiser chacune à chacune sans idée préconçue, à la recherche de corrélations. On note par exemple, dans un hôpital, pour tous les entrants, le diagnostic et les médicaments absorbés au cours des années passées, on trouve alors des associations entre certaines maladies et la prise de certains médicaments, associations qui n'avaient fait l'objet d'aucune hypothèse a priori. Dans cette démarche, la proportion des associations plausibles est quasi nulle, c'est la situation du piteux chercheur dans l'exemple des souris, et le risque vaut pratiquement 5 %. C'est trop. Comme en outre le nombre des corrélations testées est considérable, le nombre des erreurs atteint une valeur inadmissible. Ce dérapage est un sérieux danger de l'évolution actuelle. Si l'on veut tenir compte de tels résultats, il faut d'abord réduire leur nombre par un seuil de signification nettement plus sévère que 5 %, ensuite les considérer comme de simples suggestions à confirmer par des explications valables et par des enquêtes ad hoc.

Deuxième dérapage : la non-publication de résultats négatifs

Supposons qu'un phénomène n'existe pas (la drogue D n'est pas active). Sur 100 expériences, on trouvera en moyenne 5 fois qu'il existe. Si les chercheurs, ou les Revues, ne considèrent, ne publient que les résultats positifs, on verra surgir une première publication, puis d'autres — de l'ordre de 4 — apparaissant comme des confirmations, entraînant la crédibilité, ce qui ne se serait pas produit si on avait publié les 100 résultats. C'est ainsi qu'apparut un jour, dans un journal médical, la relation d'un cas de leucémie survenue chez un patient traité au long cours par la phénylbutazone. Plusieurs médecins publièrent rapidement des cas analogues. Coïncidence ou réalité? Pour le savoir, il fallait connaître les innombrables cas de patients ayant pris ce médicament, d'usage très courant, sans avoir eu ensuite de leucémie : ce qui nécessita le long travail d'un comité d'experts. La publication des résultats négatifs devrait donc être une règle. Cette règle est d'ailleurs généralement respectée pour les études portant sur des effectifs importants, elle est par contre très souvent transgressée pour les études conduites sur peu de sujets, c'est une des raisons pour n'accepter leurs conclusions qu'avec réserve.

Conclusion

Voici que se termine le temps accordé pour parler de *statistique et vérité*. Il y a surtout été question de *statistique et mensonge*. Les raisons de cette association sont d'ordre varié : le contrôle de l'erreur, qui d'une qualité fait un défaut, des dérapages liés au raisonnement déductif ou à l'emploi de chiffres, qui ne sont pas propres à la statistique mais qu'elle cautionne, enfin les dérapages propres au raisonnement statistique. Ces divers dérapages sont, dans une très large mesure, classifiables dans certaines rubriques bien précises. Qui les connaît les évite. C'est ce que fait le statisticien qualifié, nanti de sa formation et de son expérience. Parce que la statistique s'apprend. Mais cette évidence est ignorée. Chacun croit pouvoir faire sa statistique comme Monsieur Jourdain sa prose, ne suffit-il pas de savoir compter? Et chaque décideur se croit fondé à utiliser les statistiques, d'autant qu'un dérapage impercep-

tible permet de transformer une vérité déplaisante en mensonge agréable. La statistique est en cela désavantagée par rapport aux autres sciences : le premier venu oserait-il tenter une expérience de biologie moléculaire, discourir sur les conclusions d'un spécialiste, ou les utiliser? Il est certes nécessaire d'entreprendre une vulgarisation de la science statistique, mais que ce soit surtout pour développer l'esprit critique, montrer les pièges, et persuader chacun que cette science, comme toutes les autres, s'apprend.