Chiffres :
vérité ou mensonge ?

Joseph Klatzmann

Le Monde des débats, février 1995

Un État ne saurait pas plus se passer de statistiques qu'une grande entreprise ne pourrait vivre sans comptabilité. Et pour le grand public il n'est pas indifférent de savoir, par exemple, que la baisse de la natalité menace les retraites des prochaines décennies. Malheureusement, il n'est guère de statistiques qui ne posent des problèmes d'interprétation.



# Les sondages

Commençons par les sondages. Contrairement à ce qui se passe dans d'autres pays, on rappelle rarement, en France, la marge d'erreur que comportent les résultats de n'importe quelle enquête par sondage. Lorsqu'on interroge 1 000 français, on ne saurait obtenir un résultat aussi précis que si l'on en interrogeait 100 000. Lorsque cette marge atteint 4 %, cela signifie qu'un résultat égal à 52 % contre 48 % est compris entre 56 % contre 44 % et 48 % contre 52 %. Et une variation de 0,5 point en un mois, ce qu'on lit parfois, n'a évidemment aucun sens.

## Voici une conséquence de ce risque d'erreur

Supposons que la popularité réelle d'un homme politique se maintienne, d'une enquête à l'autre, à 55 %. Il pourra arriver, avec beaucoup de malchance, que tel institut de sondages trouve une chute de 59 % à 51 %, tandis qu'un autre trouvera au contraire une poussée de 51 % à 59 %. Il ne s'agit pas là d'une possibilité purement théorique, car un phénomène de ce type s'est déjà produit, à l'époque où Edith Cresson était premier ministre.

De plus, si l'erreur peut atteindre 4 % lorsqu'on interroge 1 000 personnes, elle est évidemment beaucoup plus importante lorsqu'on en interroge 100 ou moins. Apparemment, on ne fait pas d'enquêtes par sondage sur un échantillon aussi réduit. Mais cela revient au même si l'on tire d'une enquête faite auprès de 1 000 personnes des conclusions sur un groupe qui en comporte moins de 100, comme les agriculteurs ou les électeurs de tel ou tel petit parti. Or cela se fait couramment.

Encore n'ai-je parlé jusqu'à présent que de l'erreur due au fait qu'on n'interroge qu'un petit nombre de personnes.

## D'autres erreurs résultent du fait que la question est mal posée, volontairement ou non, ou de la non-sincérité de la réponse

En particulier, beaucoup de gens n'aiment pas avouer leur ignorance en répondant : « Je ne sais pas ». C'est pourquoi l'on voit 80 % des personnes interrogées émettre une opinion sur un ministre dont ils n'ont sans doute jamais entendu le nom ou dont, au mieux, ils ne savent pas de quoi il s'occupe. N'en concluez surtout pas qu'il faut rejeter systématiquement tous les résultats des enquêtes par sondage. Si des enquêtes montrent que la popularité d'un homme politique augmente de mois en mois, on peut déjà, au bout de trois mois, porter un jugement. Mais il ne faut tenir aucun compte des conclusions que certains tirent de très faibles écarts ou variations.

# L'inflation

Et les indices de prix ? Je vais ici décevoir ceux qui critiquent les statistiques avec virulence. J'ai connu l'époque où le gouvernement disposait de divers moyens (pas toujours efficaces, en raison de la fermeté de l'Insee) pour « truquer » l'indice des prix de détail. Mais lorsque l'on fait, comme aujourd'hui, un très grand nombre de relevés de prix par mois, l'indice calculé est proche de la réalité. Le gouvernement ne peut agir qu'en essayant, mais en soulevant un tollé, de supprimer de l'indice certains produits ou services dont les prix augmentent beaucoup, comme le tabac en France ou les loyers dans tel autre pays.

Des problèmes insolubles subsistent toutefois : les comparaisons dans le temps ont de moins en moins de sens, au fur et à mesure que l'on considère des périodes plus longues - en raison notamment de l'apparition de nouveaux produits et des changements de qualité des produits existants. Comment comparer le prix d'une automobile d'aujourd'hui à celui d'une automobile de 1920 ?

Une autre question est celle des erreurs de présentation des données statistiques. Que signifie pour vous la « stabilisation des dépenses médicales » ? Que celles-ci n'ont pas augmenté ? Erreur ; dans l'exemple auquel je pense, c'est le taux d'accroissement de ces dépenses qui a cessé d'augmenter. Ce type d'erreur de présentation n'est malheureusement pas rare.

Voici à quelle absurdité pourrait mener cette façon de présenter les choses. Si, dans un pays, l'inflation atteint 540 % une année et autant l'année suivante, on devrait écrire, de la même façon que ci-dessus, « stabilisation des prix » et « baisse des prix » si l'inflation revient de 540 % à 530 %. Certes, cela ne se fait pas, mais c'est le même type d'erreur que l'on fait en parlant à tort de stabilisation des dépenses médicales.

# Présentation et bon sens

Un autre type de présentation trompeuse est parfois volontaire. Pour gagner de la place sur un graphique, on ne fait pas partir les ordonnées de zéro. Si une grandeur tombe de 100 à 92 au cours d'une certaine période, cette baisse modérée apparaîtra, au premier coup d'œil, comme un effondrement si les ordonnées du graphique partent de 90 au lieu de partir de zéro.

Plus habile est la technique qui consiste à bien choisir l'année de base lorsque l'on veut calculer une évolution. Ainsi, les prix de certains produits sur le marché mondial peuvent varier considérablement d'une année à l'autre. On peut donc, à volonté, faire apparaître une hausse ou une baisse en choisissant bien la période étudiée.

Parmi les statistiques de signification trompeuse, il ne faut pas oublier les comparaisons qui concernent le produit national par habitant. On entend parler, pour les pays les plus pauvres, de produits nationaux inférieurs à 100 dollars par habitant. Dans de tels pays, les dépenses de consommation des populations les plus pauvres ne doivent même pas atteindre 50 dollars par personne. Or, avec de telles ressources, on vit certes très misérablement en Afrique, mais on ne vivrait pas du tout à Paris : avec l'équivalent de 1 franc par jour, on pourrait acheter 80 grammes de pain, fournissant quelque 200 calories - et absolument rien d'autre. Cela signifie que l'on compare des choses non comparables.

Aujourd'hui, les calculs de parités de pouvoir d'achat, même s'ils ne fournissent que des approximations grossières, sont de très loin préférables aux calculs « rigoureux » effectués en convertissant les produits nationaux en une monnaie commune quelle qu'elle soit.

# Des corrélations trompeuses

Les corrélations posent d'autres problèmes. Lorsque deux grandeurs évoluent parallèlement, sans qu'il y ait entre elles une relation rigoureuse exprimable par une équation, on dit qu'elles sont corrélées. Ainsi, il y a une corrélation positive entre la taille d'un homme et son poids - mais une corrélation seulement, car l'on peut être petit ou gros ou grand et maigre. La constatation d'une corrélation conduit à se poser deux questions : tout d'abord, y a-t-il une relation directe entre les deux phénomènes (l'un est-il la cause et l'autre l'effet ?) et puis quelle est la cause et quel est l'effet ?

Le risque d'erreur d'interprétation existe surtout pour les séries chronologiques, c'est-à-dire pour les grandeurs qui évoluent dans le temps. Certes, si l'on constate qu'au cours d'une certaine période l'espérance de vie a augmenté en même temps que s'accroissait la consommation de tabac, personne n'en tirera la conclusion que plus on fume plus on vit vieux. Mais j'ai trouvé, dans un ouvrage sérieux, un graphique montrant une évolution parallèle de la consommation de tabac et du nombre de cancers du poumon, considéré comme une preuve de la nocivité du tabac. On sait aujourd'hui avec certitude que la consommation de tabac peut provoquer le cancer du poumon, mais ce n'est pas le graphique cité ci-dessus qui le prouve. Par ailleurs, dans le cas de la corrélation entre consommation de tabac et cancer du poumon, il n'y a aucun doute en ce qui concerne la cause et l'effet. Mais si l'on trouve une corrélation positive entre la pratique d'un exercice physique et l'état du système cardio-vasculaire, on en tire automatiquement la conclusion que c'est l'exercice qui agit favorablement sur le cœur, alors que la corrélation inverse n'a rien d'invraisemblable. Il n'est en effet pas impossible que ceux qui ont le cœur solide - qu'ils le sachent ou non - soient spontanément incités à pratiquer un exercice physique.

Toujours en ce qui concerne les corrélations, on peut trouver n'importe quoi en ne prenant en compte qu'un petit nombre de cas. Ainsi, si l'on étudie la relation entre nombre de médecins pour 100 000 habitants et espérance de vie, on trouve les résultats suivants en considérant la France, la Russie et le Japon : c'est en Russie que le nombre de médecins est le plus élevé et l'espérance de vie la plus courte, tandis que le Japon a à la fois la plus faible proportion de médecins et la durée de vie la plus longue. Conclusion « évidente » : moins il y a de médecins et plus on vit vieux ! Personne n'a tiré une telle conclusion. C'est vrai, mais un auteur français qui nie l'utilité de la médecine s'est servi sans la discuter d'une information relative à une diminution de la mortalité en Israël pendant une longue grève des médecins pour y trouver une confirmation de sa thèse. On pourrait multiplier les exemples sans fin ; rien ne serait plus facile. Je pourrais citer encore la vogue des chiffres totalement dénués de sens parce que leur signification dépend de la population du pays : un chômeur de plus toutes les deux minutes, cela serait infiniment plus grave pour le Luxembourg que pour la Chine et, s'il s'agit de la France, on ne peut, sans se livrer à des calculs, savoir si c'est beaucoup ou peu ; les affirmations sans preuves et les chiffres totalement aberrants (comme les 50 millions de morts de faim par an), etc.

# En un mot : réfléchissez

Inutile de continuer. La conclusion s'impose : il faut se méfier de la plupart des statistiques. Il faut savoir résister à deux tentations : rejeter tout en bloc (« ce ne sont que des statistiques ») ou accepter tout sans réfléchir. Cette seconde tentation est plus dangereuse qu'on ne croit, car il est parfois très difficile, même à une personne bien avertie, de ne pas y succomber : on cherche un chiffre difficile à obtenir ; si on le trouve, on est si content que l'on oublie de se poser des questions, ou bien le chiffre trouvé confirme ce que l'on pensait, et tout esprit critique disparaît.

Que faire alors, comment se comporter en face des données statistiques, surtout si l'on n'est pas parfaitement au courant de tous les pièges que celles-ci recèlent ?