



**Mathématiques
et sciences humaines**
Mathematics and social sciences

153 | 2001
Varia

L'interprétation des résultats statistiques

Interpreting statistical results

Thierry Foucart



Édition électronique

URL : <http://msh.revues.org/2840>
DOI : 10.4000/msh.2840
ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique
sociales de l'EHESS

Édition imprimée

Date de publication : 1 mars 2001
ISSN : 0987-6936

Référence électronique

Thierry Foucart, « L'interprétation des résultats statistiques », *Mathématiques et sciences humaines* [En ligne], 153 | Printemps 2001, mis en ligne le 10 février 2006, consulté le 30 septembre 2016. URL : <http://msh.revues.org/2840> ; DOI : 10.4000/msh.2840

Ce document est un fac-similé de l'édition imprimée.

© École des hautes études en sciences sociales

L'INTERPRÉTATION DES RÉSULTATS STATISTIQUES

Thierry FOUCART¹

RÉSUMÉ — *La plupart des mesures sociales sont fondées actuellement sur des résultats statistiques et leurs interprétations. Cette argumentation est parfois erronée, par manque de compétence et d'esprit critique, et abusive lorsque les chiffres et leurs interprétations servent à justifier des idées préconçues ou des motivations personnelles conscientes ou non. Nous proposons dans cet article, une mise en garde contre ces interprétations destinée à la fois à leurs auteurs et ceux qui les lisent.*

MOTS-CLÉS — Corrélation, Causalité, Chance, Modélisation, Sciences sociales.

SUMMARY — Interpreting statistical results.

Most of the social measures are founded currently on statistical results and their interpretations. This argumentation is sometimes erroneous, for lack of competence and critical spirit, and abusive when the figures and their interpretations are used to justify preconceived ideas or personal motivations which can be conscious or not. The intention of this article is to warn people, both authors and readers, against these interpretations.

KEYWORDS — Correlation, Causality, Chance, Modelling, Social sciences.

INTRODUCTION

La statistique appliquée consiste à chercher de l'information dans des données nombreuses, obtenues par enquête, sondage, tirage au hasard ou raisonné, à l'aide de ce que l'on appelle un modèle, c'est-à-dire une « représentation mathématique d'un phénomène physique, économique, humain, etc., réalisée afin de pouvoir mieux étudier celui-ci ». (Petit Larousse).

Les développements scientifiques récents ont permis une modélisation de plus en plus complexe et une application de plus en plus facile : la statistique appliquée est ainsi devenue l'un des fondements de la démarche scientifique dans toutes les sciences appliquées, physique, chimie, médecine, gestion, mais aussi en sociologie, en psychologie et dans les sciences de l'éducation.

LES CONSÉQUENCES DE LA VULGARISATION STATISTIQUE

La facilité avec laquelle on peut procéder maintenant à une enquête, à une analyse de données, met la modélisation et la statistique à la portée d'un grand nombre d'utilisateurs

¹ Agrégé de mathématiques, Maître de conférences à l'université de Poitiers, e-mail : tfoucart@wanadoo.fr.

qui n'ont pas nécessairement les connaissances en statistique suffisantes pour en appliquer correctement les méthodes. Paradoxalement, l'augmentation considérable de la quantité d'informations statistiques que l'informatique permet d'obtenir a pour conséquence d'accroître l'exigence dans la compétence scientifique et l'esprit critique des utilisateurs. On est actuellement très loin de répondre à cette exigence : il ne suffit pas d'être mathématicien, informaticien, médecin, gestionnaire, sociologue ou psychologue pour mettre en œuvre une méthode statistique et en analyser correctement les résultats. Il ne suffit pas non plus d'être statisticien □ il faut posséder des compétences multiples, ou travailler en équipe pluridisciplinaire.

Une approche méthodologique correcte au plan statistique donne en général des informations précises comme on peut le vérifier en comparant les résultats des sondages préélectoraux aux résultats des élections. Mais inversement, une enquête mal menée, un questionnaire mal rédigé ou un échantillon mal choisi peuvent aboutir à des résultats statistiques complètement erronés. L'enquête effectuée auprès des enseignants et des lycéens en 1997, comme celle de 1994, n'a, par exemple, aucune valeur statistique malgré le grand nombre de questionnaires recueillis : comme le dit lui-même Philippe Meirieu, responsable de la consultation de 1997, « nous ne pouvons pas prétendre à une approche scientifique »¹.

Ce problème de compétence se double chez l'utilisateur et dans le public d'une impression de rigueur, d'exactitude dans les résultats quand ils sont présentés de manière chiffrée, plus ou moins scientifique. Cette présentation augmente la confiance que l'on accorde aux résultats numériques fondés sur une modélisation numérique, alors que leur interprétation demande toujours une aussi grande prudence. Un résultat exprimé sous forme de pourcentage (74,9 % par exemple) crée l'illusion d'une plus grande précision qu'un résultat exprimé de façon courante (une forte majorité) alors qu'ils fournissent tous deux la même information.

En outre, les pourcentages sont interprétés souvent comme une probabilité, une chance □ « Des travaux récents de l'Insee montrent que pour les titulaires du baccalauréat général, au bout de 10 ans d'expérience, les chances d'occuper un emploi de cadre sont de 17 % pour les hommes et de 8 % pour les femmes »². Ce n'est pas une *chance* de devenir cadre □ c'est une démarche individuelle, sans modèle probabiliste ni tirage au hasard. Le vocabulaire employé par Margaret Maruani et l'utilisation des pourcentages reviennent à comparer la *chance* qu'a chacun de devenir cadre avec celle qu'il a de gagner au loto, ce qui n'a évidemment aucun sens.

LES DIFFICULTÉS DE L'INTERPRÉTATION STATISTIQUE

En sociologie et psychologie, la présentation chiffrée est issue d'un codage informatique de situations et de comportements humains qui n'en donne qu'une information partielle puisque quantifiable. L'information contenue dans ces codages n'est pas du tout exhaustive, ignore le non-dit, l'intuitif, le sentiment.

¹ « Le casse-tête du dépouillement et de l'analyse des réponses », Sandrine Blanchard, *Le Monde*, 24 janvier 1998.

² Margaret Maruani, *La Documentation Française*, n° 291, mai-juin 1999.

On ne dénombre souvent que des déclarations□ce n'est pas du tout la même chose que le fait lui-même, surtout lorsqu'il s'agit d'études de comportements. La difficulté ne se limite pas au recueil des informations, elle concerne aussi le sens de l'information donnée.

« La difficulté est que, selon les cas, l'observation "une majorité pense que X" peut être prise à sa valeur faciale et constituer une information essentielle pour le politique, alors que, dans d'autres cas, elle peut à l'extrême refléter des intérêts individuels myopes. Il est donc important d'apprendre à lire les données d'opinion en fonction des motivations des opinants » ¹.

On pourrait ajouter une réflexion analogue concernant ceux qui interprètent les données, certains, parmi eux, lisant les données d'opinion en fonction de leurs motivations personnelles. En effet, les sociologues considèrent qu'«Une relation statistique n'a généralement de sens que si elle est interprétable en termes de causalité»². C'est dans la recherche d'une telle interprétation que les motivations personnelles du chercheur, sociologue, psychologue ou autre, sont très présentes.

La notion de causalité n'a pas de signification statistique précise. Détectée sur un ensemble d'observations, une relation statistique ne peut concerner que cet ensemble et non chacune des observations.

Prenons un exemple précis : le coefficient de corrélation entre la taille et le poids, calculé sur 90 jeunes filles de 20 ans, est égal à 0.3991³. L'analyse statistique montre l'existence d'une relation linéaire entre les deux variables, les jeunes filles ayant généralement une taille et un poids simultanément supérieurs à la moyenne, ou simultanément inférieurs. Mais□

- cette relation n'est pas vérifiée par environ 35 % des observations□
- rien ne permet d'affirmer que cette relation est due, dans les 65 % d'observations restantes, à une même cause.

La causalité ne peut donc s'interpréter que d'une façon collective, et interpréter une relation statistique en terme de causalité signifie□précisément : *c'est parce qu'un groupe possède majoritairement telle propriété qu'il possède majoritairement telle autre*. Mais cela n'a pas de rapport avec une causalité individuelle bien difficile à détecter statistiquement.

D'ailleurs, cela n'a pas de rapport non plus avec une relation causale concernant des sous-groupes□il est très possible qu'une relation significative entre deux variables existant sur une population d'individus disparaisse ou même soit inversée lorsqu'on observe un sous-groupe de cette population.

Inversement, une relation de causalité peut se traduire par un coefficient de corrélation nul ou même de signe inverse□le trafic annuel routier et le nombre annuel d'accidents sur les routes ont ainsi un coefficient de corrélation de -0.98, alors

¹ Raymond Boudon, «Du bon usage des sondages en politique», *Commentaire*, n° 93, printemps 2001.

² Raymond Boudon, *Encyclopaedia Universalis*, article *Causalité*.

³ On trouvera ces données dans *L'analyse des données mode d'emploi*, T. Foucart, Presses universitaires de Rennes, 1997.

qu'évidemment l'augmentation du nombre de véhicules sur les routes devrait provoquer plus d'accidents□ c'est la simultanéité de la décroissance du nombre d'accidents et de l'augmentation du trafic qui explique cette valeur.

En outre, la causalité ne peut être que relative, de la même façon qu'un accident de la route est dû plutôt à un enchaînement de circonstances qu'à une cause unique. Le conducteur allait trop vite, il était fatigué, il y avait un virage, un camion roulait à gauche. Si nous supprimons une seule de ces conditions, il n'y aurait pas eu d'accident□ il n'y a pas une seule cause, mais plusieurs.

La causalité demande donc une description de la réalité à laquelle on se limite. Les interprétations d'une relation statistique ne sont pas des vérités objectives en ce sens que deux personnes peuvent en proposer des interprétations différentes puisqu'elle est interprétée nécessairement dans un contexte différent, ne serait-ce qu'à cause des personnalités différentes.

L'HYPOTHÈSE «TOUTES CHOSES ÉGALES PAR AILLEURS»

Pour isoler un facteur parmi plusieurs et en examiner l'importance dans la relation entre deux phénomènes, on est amené à contrôler la réalité au sein de laquelle on effectue les observations□ on raisonne alors «toutes choses égales par ailleurs». Par exemple, pour comparer les salaires entre les hommes et les femmes, on compare les salaires de personnes qui ne diffèrent que par le sexe parmi les facteurs explicatifs de la rémunération du travail□ «toutes choses égales par ailleurs, c'est-à-dire à niveaux de formation, d'expérience, de catégorie socioprofessionnelle et d'âge équivalents, il reste un écart de 12 % [N.B. de salaire au bénéfice des hommes]»¹.

La relation statistique semble claire□ les femmes gagnent moins que les hommes «toutes choses égales par ailleurs». Dès lors, certains sociologues interprètent la relation statistique comme une relation causale et affirment que la société dévalue le travail des femmes parce que ce sont des femmes.

Cette démarche est subjective□ on choisit pour différencier les gens un critère particulier – le sexe –, en le considérant de facto comme facteur explicatif, mais on aurait pu tout autant choisir la race, la religion, l'âge, ... Des analyses suivant ces critères auraient très vraisemblablement abouti à des conclusions analogues, ou à des conclusions inverses dangereuses². En effet, l'argument se retourne□ le fait que les femmes soient moins bien rémunérées ne montre-t-il pas que le travail fourni n'est pas le même, puisque la loi impose l'égalité entre hommes et femmes□

On peut contester cette condition «toutes choses égales par ailleurs», en particulier le choix des facteurs explicatifs qui doivent être complets et sans redondance□ c'est une difficulté bien connue du choix des variables explicatives dans le modèle linéaire multiple.

¹ Margaret Maruani, op. cit.

² Comme dans l'ouvrage *The Bell Curve* de C. Murray et R. Herrnstein qui concluent à une échelle raciale en comparant les Blancs, les Jaunes et les Noirs aux États-Unis.

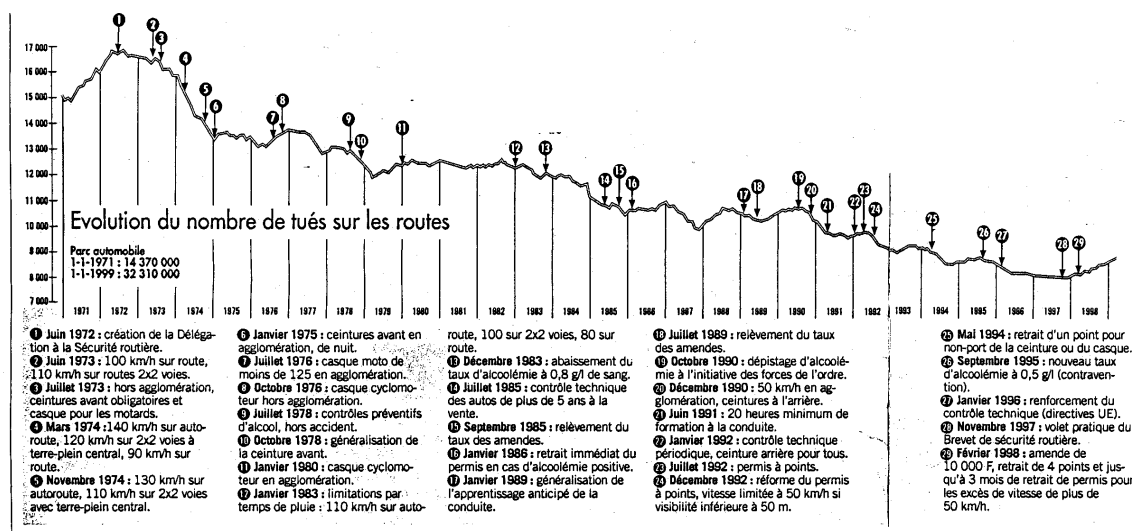
Dans d'autres enquêtes, c'est effectivement cette condition qui est remise en cause, éventuellement par les mêmes sociologues. Ainsi, le fait que les filles travaillent mieux que les garçons au collège et au lycée est expliquée par l'éducation différente qui leur est donnée. On ne déduit pas de l'inégalité des résultats que le système éducatif est injuste vis-à-vis des garçons. Mais la différence reconnue d'éducation entre les garçons et les filles n'est pas considérée comme facteur explicatif du travail et n'est donc pas incluse dans l'hypothèse «toutes choses égales par ailleurs» émise à propos de l'inégalité des salaires hommes-femmes.

L'expérience montre que plus on tient compte de facteurs dans cette condition : âge, diplôme, secteur d'activité, localisation, entreprise... , plus le nombre de personnes comparables diminue, et plus il est difficile d'en tirer des conclusions. En réalité, la condition «toutes choses égales par ailleurs» ne peut jamais être totalement réalisée, parfois même par contradiction interne. Comment faire pour comparer le salaire d'un homme à celui d'une femme «toutes choses égales par ailleurs», puisque dans le premier cas, le conjoint est une femme, et dans le second, c'est un homme. Cette condition est finalement une hypothèse définie par le sociologue et vérifiée approximativement, dont les conséquences ne peuvent être que des suppositions émises inévitablement en fonction de la personnalité de leur auteur.

L'ANALYSE DES GRAPHIQUES

Les graphiques sont aussi fréquemment utilisés pour mettre en évidence une relation entre deux séries de données.

Le graphique ci-dessous donne l'évolution du nombre de tués sur les routes françaises et les mesures de sécurité routières prises au cours de la période étudiée (revue *Automoto*, août 99).



Une interprétation superficielle laisse croire que ce graphique prouve que le nombre de tués a diminué *grâce aux* mesures prises. Mais c'est faux : ce n'est pas une preuve. On aurait pu tout aussi bien indiquer les résultats des élections australiennes à la place des mesures de sécurité et (ou) le taux mensuel d'inflation à la place du nombre de tués. Peut-être constaterait-on alors la coïncidence des évolutions sans que l'on puisse expliquer l'une par l'autre ?

Pour prouver la relation de causalité entre la diminution du nombre de morts et une mesure de prévention routière, il faudrait vérifier l'impact de cette dernière sur les accidents survenus auparavant : c'est en examinant les circonstances dans lesquelles ils se sont produits que l'on peut évaluer l'efficacité de la mesure prise.

On peut donner un autre exemple de ce type de raisonnement erroné : une société publie dans une revue financière un graphique montrant la hausse du cours d'un fonds financier international de 1935 à nos jours, en indiquant, comme précédemment pour un certain nombre de dates, des événements caractéristiques de la période : il s'agit ici d'événements politiques tels que les guerres de 39-45, de Corée, d'Indochine, d'Algérie, du Vietnam... allant à l'évidence à l'encontre du développement économique.

L'objectif est visiblement de montrer que la gestion du fonds a permis une croissance quasi ininterrompue de la valeur financière de la part *malgré* tous ces événements. Mais que se serait-il passé *sans* tous ces événements ? La démarche correcte est ici aussi de montrer que les choix d'investissement, à la suite de chaque événement, ont été effectués de façon judicieuse.

L'exploitation à des fins publicitaires de ce dernier graphique consiste finalement à inverser le raisonnement par rapport au précédent, l'expression *grâce à* étant changée par *malgré*. Ces deux exemples contradictoires dans la démarche prouvent bien que les graphiques sont présentés pour convaincre et non pour démontrer, et qu'il faut mener une réflexion critique sur la façon dont ils sont construits avant d'admettre ce qu'ils prétendent prouver.

LES LIMITES DE LA MODÉLISATION

Pour représenter correctement un phénomène à l'aide de formules mathématiques, qu'il concerne la réalité physique, sociale ou psychologique, on est amené à effectuer un certain nombre d'hypothèses.

Voici par exemple un modèle pour caractériser l'intelligence tel que le présente Yela Mariano dans l'*Encyclopaedia Universalis* (article *Analyse factorielle*) :

$$\begin{array}{ll} \ll & (1) \quad j = f(g, s) \\ & (2) \quad r_{gs} = r_{sj} = 0 \\ & (3) \quad z_j = a_{jg} + b_{js} \end{array}$$

Ces égalités signifient : qu'une activité cognitive quelconque j est fonction d'un facteur général, g , commun à toutes les activités, et d'un facteur spécifique, s , lequel n'est présent que dans cette activité ; que tous les facteurs sont statistiquement indépendants, c'est-à-dire que leurs corrélations sont nulles ; que la mesure z_j d'un comportement j peut être écrite en première approximation comme une fonction linéaire de g et de s ; c'est-à-dire qu'elle est composée d'une partie due

au facteur g représentée par le coefficient factoriel de j en g a_{jg} , augmentée d'une part due au facteur spécifique s représentée par le coefficient factoriel de j en s b_{js} (3) »

Ce modèle est la formalisation mathématique de la démarche explicative des activités cognitives imaginée par Spearman. Il est à l'origine de l'analyse unifactorielle, parce qu'il n'envisage qu'un facteur g , généralisée par la suite en analyse multifactorielle (ou simplement factorielle) qui prend en compte plusieurs facteurs généraux éventuels.

On ne connaît pas *a priori* le facteur g ni les facteurs spécifiques s , et l'objectif de l'analyse est de les mettre en évidence à l'aide d'expériences soigneusement choisies, effectuées sur un certain nombre de personnes. Il s'agit donc de valider la structure de l'intelligence imaginée par Spearman.

Il y a trois points fondamentaux à respecter pour que l'expérimentation possède un caractère scientifique :

- les activités cognitives choisies doivent être suffisamment caractéristiques des facteurs, dont l'ensemble représente ce que l'on peut appeler l'intelligence—cela peut être des tests de mémoire, de compréhension, d'invention—
- les personnes soumises à ces tests sont tirées au hasard dans la population concernée—on dispose ainsi des outils de la statistique inférentielle qui permettent d'évaluer la part du hasard dans les résultats (intervalles de confiance, tests statistiques)—
- pour vérifier le modèle estimé, on contrôle les hypothèses traduites par le modèle ci-dessus, si possible sur un échantillon constitué d'autres personnes soumises aux mêmes tests et tirées au hasard dans la même population.

Les deux premiers points sont relativement faciles à respecter, mais la vérification complète du modèle est impossible—la statistique ne permet jamais de vérifier qu'une hypothèse est vraie, seulement qu'elle est vraisemblable. Accepter la nullité d'un coefficient de corrélation théorique signifie simplement que les observations effectuées ne sont pas en contradiction avec cette hypothèse—le test contrôle le risque de première espèce, mais pas le risque de seconde espèce¹. En outre, la nullité d'un coefficient de corrélation n'implique l'indépendance des variables que si les lois mises en jeu sont gaussiennes, ce qui n'est jamais établi avec certitude. La validation du modèle ne consiste finalement qu'à montrer qu'il n'est pas intrinsèquement contradictoire.

Imaginons maintenant que le modèle de Spearman fasse référence à un facteur général g et quatre facteurs spécifiques s . Il faut donc supposer l'indépendance de cinq facteurs deux à deux, ce qui revient à effectuer dix hypothèses d'indépendance. En accordant un degré de confiance de 90 % à chacune d'entre elles, on peut calculer le degré de confiance de l'ensemble, qui est de l'ordre de 35 % ($= 0,9^{10}$).

Tout cela incite à la prudence lorsque l'on considère le modèle satisfaisant dans son ensemble, et montre l'importance d'une justification théorique des hypothèses, que

¹ Le risque de première espèce est ici la probabilité de rejeter la nullité du coefficient de corrélation alors qu'elle est vraie, le risque de seconde espèce, la probabilité d'accepter cette nullité alors qu'elle est fausse.

l'on peut donner parfois dans les sciences exactes, au lieu d'une simple vérification *a posteriori*, seule procédure possible dans les sciences sociales.

CONCLUSION

La statistique appliquée ne donne qu'une image approximative de la réalité qui nous entoure, beaucoup trop complexe pour être contenue dans une liste de nombres aussi grande soit-elle. La modélisation n'est qu'un outil supplémentaire d'observation, et ne peut représenter un phénomène dans sa globalité.

L'analyse des résultats doit être menée avec un esprit critique attentif, une compétence certaine dans le champ scientifique étudié et une honnêteté intellectuelle incontestable.

C'est l'explication scientifique des coïncidences mises en évidence qui doit permettre de prendre des décisions politiques et sociales, et non la coïncidence elle-même. Nous avons le sentiment que cette démarche est très souvent oubliée à l'heure actuelle.