

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

PIERRE DAGNELIE

Diversité et unité de la statistique

Journal de la société statistique de Paris, tome 123, n° 2 (1982), p. 86-92

http://www.numdam.org/item?id=JSFS_1982__123_2_86_0

© Société de statistique de Paris, 1982, 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/>

I — COMMUNICATIONS

DIVERSITÉ ET UNITÉ DE LA STATISTIQUE (1)

Pierre DAGNELIE

Professeur de statistique à la Faculté des sciences agronomiques de GEMBLoux (Belgique)

L'auteur met en évidence différentes sources de diversité de la statistique : orientation générale des travaux, domaines d'application, niveau de développement des pays, langues, etc. Il souligne cependant les liens qui existent entre ces différentes tendances et insiste en conclusion sur le fait que cette diversité peut être particulièrement féconde dans le cadre de contacts et d'échanges entre statisticiens d'orientations différentes.

The author shows various sources of diversity in statistics : general orientation of works, application field, development level of countries, languages, etc. However, he underlines the links between these various trends and insists, in conclusion, on how fruitful this diversity in contacts and exchanges may be between statisticians of various orientations.

Der Verfasser untersucht die verschiedenen Quellen verschiedener Statistiken : Die allgemeinen Basen der verschiedenen Statistiken, die Gebiete ihrer Anwendung, das Entwicklungsniveau der einzelnen Länder, die Sprachen, etc. Er unterstreicht die Verbindungen, die zwischen den verschiedenen Tendenzen bestehen und besteht in seinen Schlussfolgerungen auf der Tatsache, dass diese Verschiedenheiten besonders fruchtbar im Rahmen der Kontakte und dem Austausch von Ideen unter Statistikern verschiedener Richtungen sein können.

A l'occasion de cette remise de prix, il m'a paru opportun de présenter à la Société de Statistique de Paris et de France une communication d'intérêt général, sous la forme de quelques réflexions relatives à la diversité et l'unité de la statistique.

Dans ce but, je pourrais tout d'abord consacrer un certain temps à définir ou à re-définir ce que l'on entend par statistique, en retournant par exemple aux quelque 100 ou 120 définitions réunies par WILLCOX en 1935 ou à la discussion dont vous avez publié le compte rendu en 1956 (2). Tel n'est cependant pas mon objectif et je me contenterai de rappeler ici les deux acceptions les plus classiques, je pense, auxquelles je me réfère d'ailleurs habituellement : d'une part, *la statistique ou les statistiques*, au singulier ou au pluriel, désignant des ensembles cohérents de données numériques relatives à des groupes d'individus, au sens large, et d'autre part, *la statistique*, considérée

Journal de la Société de statistique de Paris, tome 123, n° 2, 1982.

(1) Communication présentée le 22 janvier 1982 devant les Sociétés de statistique de Paris et de France, lors de la remise du prix du statisticien d'expression française 1981.

(2) WILLCOX W.F. [1935]. Definitions of statistics. *Rev. Inst. Int. Stat.* 3, 388-399.

Discussion sur la définition du mot « statistique ». *J. Soc. Stat. Paris* 97, n° 10-11-12, 1956, 253-258.

comme l'ensemble des méthodes qui permettent de réunir et d'analyser de telles données (1).

C'est en fait de cette statistique-là, de cet ensemble de méthodes, que je souhaite parler ici.

**

Pour illustrer la DIVERSITÉ DE LA STATISTIQUE, je voudrais, dans un premier temps et de façon certes schématique, tenter d'identifier *trois grandes tendances*.

Tout d'abord, la statistique que je qualifierai d'« administrative » ou « gouvernementale », c'est-à-dire celle qui se fait dans les instituts nationaux et internationaux de statistique, tels que l'Institut National (français) de la Statistique et des Études Économiques, à Paris, ou l'Office Statistique des Communautés Européennes, à Luxembourg. Il s'agit d'une statistique relative généralement à de grands ensembles de données, réunies le plus souvent par enquêtes, et dans une large mesure aussi, d'une statistique relativement traditionnelle, quant aux méthodes utilisées, et cela en raison précisément de la continuité du travail qui doit être réalisé dans ces instituts.

En deuxième lieu, la statistique dite « mathématique », que l'on pourrait aussi qualifier d'« universitaire ». C'est la statistique qui se développe habituellement dans les instituts de mathématique des universités, souvent avec bien peu de données numériques, voire sans aucune donnée numérique ou sans données numériques réelles; avec une volonté de novation, mais une volonté de novation qui reste, ou qui est restée fréquemment limitée au seul cadre gaussien; avec aussi des frontières relativement floues par rapport au calcul ou à la théorie des probabilités.

D'autre part encore, la statistique « appliquée » ou « de terrain », c'est-à-dire celle qui se pratique dans les instituts de sondage d'opinions, dans certaines grandes écoles, dans de nombreuses institutions de recherche appliquée ou orientée, dans des facultés de médecine, etc.; avec une approche assez différente, dont le principe essentiel est en général de partir des problèmes concrets, et de ce fait aussi, avec une diversité souvent plus importante.

Sans doute, ce découpage en trois types de statistique peut-il paraître quelque peu outrancier. La statistique que j'ai qualifiée d'administrative est devenue de moins en moins traditionnelle au fil du temps, avec une utilisation plus fréquente de méthodes nouvelles; de même aussi, la statistique universitaire, au moins dans certaines institutions, aborde plus souvent que par le passé le traitement de données numériques réelles; de même enfin, la statistique que j'ai dite appliquée devient parfois très mathématique, voire même, me semble-t-il, abusivement mathématique. Mais cependant, je pense que les trois tendances que j'ai citées caractérisent assez bien la statistique actuelle, telle que nous pouvons la voir tous les jours.

**

Une deuxième distinction dont il faut parler, surtout dans le cadre de la statistique appliquée, est celle qui est liée aux *différents domaines d'application*.

(1) DAGNELIE P. [1979-1980]. *Théorie et méthodes statistiques* (2 vol.). Gembloux, Presses Agronomiques, 378 + 463 p.

La biométrie et la psychométrie sont des notions qui ont plus de 80 ans d'âge et qui sont bien connues de tous, pour désigner l'application des méthodes quantitatives, essentiellement statistiques, aux sciences biologiques et psychologiques. De même, le mot économétrie a déjà, lui également, une longue histoire. Mais d'autres vocables semblables sont apparus plus récemment et d'autres domaines d'application importants de la statistique sont nés, sans être baptisés cependant de la même façon : tel est le cas par exemple de l'archéométrie, de la statistique appliquée aux sciences géographiques (qui n'est pas de la « géométrie »!), des applications linguistiques de la statistique, etc.

A un niveau inférieur, d'autres appellations sont aussi utilisées plus ou moins couramment. Ainsi, dans le cadre de la biométrie, on parlera de la dendrométrie pour désigner l'étude quantitative des arbres et parfois aussi de la dasométrie pour désigner l'étude quantitative des peuplements forestiers.



Une autre distinction encore est celle que l'on peut faire entre *la statistique du riche et la statistique du pauvre*.

Je veux faire allusion tout d'abord à la statistique des pays riches, par opposition à celle des pays en voie de développement. En fait, nous sommes habitués à travailler avec tous les moyens qui sont à notre disposition et nous oublions trop souvent que, dans beaucoup de pays et pour une très large part de l'humanité, la statistique se fait encore sans ordinateurs, sans moyens de calcul importants, sans documentation ou sans guère de documentation, et avec cependant tous nos problèmes classiques et toutes nos difficultés courantes, auxquels s'ajoutent généralement des problèmes spécifiques qui ne font que compliquer les choses.

Non seulement nous avons trop tendance à oublier l'existence même de cette disparité, mais en outre, nous n'en mesurons sans doute pas toute l'importance et nous ne réalisons pas quelle est son évolution. En vue de la prochaine Conférence Internationale de Biométrie, qui doit se tenir à Toulouse et au cours de laquelle je dois présider une réunion relative à la situation de la biométrie dans les pays du Tiers Monde, j'ai feuilleté les annuaires de la Société Internationale de Biométrie, en dénombrant les membres de cette Société appartenant aux pays riches d'une part et aux pays en voie de développement d'autre part et j'ai comparé ces nombres de membres aux populations totales des deux catégories de pays (1). Les résultats sont édifiants.

Très rapidement, on peut dire qu'en 1957, 10 % des membres de la Société Internationale de Biométrie appartenaient aux pays du Tiers Monde et que, 25 ans plus tard, en 1982, cette proportion est vraisemblablement de l'ordre de 6 %. En 1957, proportionnellement aux nombres d'habitants, il y avait environ 20 fois plus de membres de la Société Internationale de Biométrie dans les pays développés que dans les pays du Tiers Monde; 25 ans plus tard à nouveau, en 1982, et toujours proportionnellement aux nombres d'habitants, il y a environ 45 fois plus de membres dans les pays développés que dans les pays du Tiers Monde (2). Ce qui m'amène à penser que la biométrie des pays

(1) DAGNELIE P. [1982]. *La situation de la biométrie dans les pays du Tiers Monde*. Toulouse, Conf. Intern. Biom. (en préparation).

(2) En 1957 : 150 membres par 100 000 000 habitants dans les pays développés et 8 membres pour 100 000 000 habitants dans les pays du Tiers Monde; en 1982, vraisemblablement 360 membres par 100 000 000 habitants dans les pays développés et toujours 8 membres par 100 000 000 habitants dans les pays du Tiers Monde.

pauvres est vraisemblablement deux à trois fois plus pauvre maintenant qu'il y a 25 ans, toutes proportions gardées, par rapport à la population totale.

Bien sûr, la biométrie n'est pas la statistique, l'affiliation à une société internationale n'est qu'un indice et ces chiffres ne sont qu'approximatifs. Mais je crois sincèrement que les tendances qu'ils mettent en évidence correspondent bien à la réalité.

Indépendamment de cette disparité considérable, et croissante, entre pays riches et pays pauvres, il existe aussi des disparités non négligeables aux niveaux national, régional et local. Dans nos pays d'Europe occidentale, je pense par exemple aux différences de conditions de travail qui séparent le statisticien d'un grand centre universitaire et le statisticien isolé œuvrant dans une petite ville de province ou dans une entreprise de faible dimension. Dans certains grands pays en rapide expansion, je pense aux différences entre les régions les plus industrialisées et les régions les plus reculées : le sud-est du Brésil, par exemple, par opposition à l'Amazonie et au Nordeste. Enfin, je ne peux m'empêcher de penser aussi aux pays les plus pauvres, à certains pays d'Afrique notamment, où la recherche agronomique, donc alimentaire, se réalise encore, y compris l'analyse statistique des résultats, dans des stations qui bien sûr ne possèdent ni ordinateurs, ni machines à calculer importantes, mais qui n'ont même pas d'électricité, ni de raccordement au réseau téléphonique, et donc aucune possibilité de communication rapide avec l'extérieur.

*
**

« Et l'*analyse des données*? », me diront certains, « dans tout cela, qu'en faites-vous? ».

Je n'en parlerai pas longuement à l'occasion de cette réunion, car j'ai déjà eu la possibilité d'exprimer mon point de vue à ce sujet lors des Premières Journées « Analyse des Données et Informatique », qui ont été organisées à Versailles en 1977. Sans sous-estimer aucunement l'apport récent de l'école française, j'ai en fait le sentiment que l'analyse des données doit être considérée comme faisant partie intégrante de la statistique, et plus particulièrement de l'analyse statistique à plusieurs variables ⁽¹⁾.

Au risque de paraître peut-être sévère, je voudrais ajouter ici que, par comparaison avec les disparités auxquelles je viens de faire allusion, les divergences entre l'analyse des données française et l'analyse multivariée anglo-saxonne me semblent être plus une querelle de riches qu'une réalité.

*
**

Enfin, une dernière diversité est celle des *langues*: des langues officielles tout d'abord, comme le français et l'anglais, mais aussi des jargons propres notamment aux différentes écoles, aux différents domaines d'application de la statistique, etc.

Le statisticien qui côtoie des collègues de disciplines diverses se rend compte en effet que des notions parfois très semblables peuvent être désignées par des vocables très différents. La notion de coefficient d'allométrie du biométricien, par exemple, n'est pas

⁽¹⁾ DAGNELIE P. [1977]. Quelques réflexions relatives à l'analyse des données. In : *Premières Journées Internationales « Analyse des Données et Informatique »*. Le Chesnay, Institut de Recherche d'Informatique et d'Automatique, 353-363.

fort éloignée de la notion d'élasticité de l'économètre (et non de l'économétricien : encore une nuance de notre langue!).

Et il faut reconnaître que le développement récent de la statistique n'a pas simplifié les choses. Pouvons-nous croire, par exemple, qu'il n'y ait aucune confusion, même dans l'esprit de certains statisticiens chevronnés, entre analyse factorielle, analyse d'une expérience factorielle, analyse factorielle des correspondances, analyse factorielle discriminante, analyse discriminante, analyse canonique, analyse des corrélations canoniques, etc? Et la liste pourrait être considérablement allongée.



Je voudrais conclure cette première partie de mon exposé en constatant que les institutions, la géographie, les langues et surtout l'homme, tout concourt à la *diversité de la statistique*. Et je pense que cette diversité doit être considérée dans une très large mesure comme un élément positif, car elle est la source d'une réelle *richesse collective*.

Mais toute pièce a son revers et il faut bien reconnaître que certains, dans notre profession, se replient volontiers sur eux-mêmes, restent confinés dans leurs propres préoccupations et utilisent, consciemment ou non, un vocabulaire chaque jour plus ésotérique, comme si le fait de ne pas être compris ou d'être mal compris permettait de mieux affirmer sa supériorité. Dans cette optique, il faudrait dire que les institutions, la géographie, les langues et l'homme, tout concourt ou, à tout le moins, tout pourrait concourir, non pas à la diversité, mais à la *division de la statistique*. Et cela dans un sens très défavorable, car une telle division est inévitablement la source d'une relative *pauvreté individuelle* de chacun.



Dans la deuxième partie de mon exposé, je voudrais, plus rapidement, parler de l'UNITÉ DE LA STATISTIQUE, en montrant par quelques exemples ce que chaque forme ou chaque type de statistique que j'ai esquissé a apporté et peut apporter aux autres.

Au niveau des *trois tendances générales* que j'ai essayé d'identifier en commençant, je ne crois pas qu'il soit nécessaire de mettre en évidence l'apport, en méthodes nouvelles, de la statistique mathématique ou universitaire à la statistique administrative ou gouvernementale et à la statistique appliquée. Cet apport de la théorie à la pratique est évidemment tout à fait naturel.

Il est intéressant par contre de souligner le rôle de méthodologies telles que la réalisation d'enquêtes par échantillonnage, l'étude des modèles économétriques ou équations de régression simultanées, la planification des expériences et l'analyse de leurs résultats. Ces méthodologies, qui se sont développées au départ, respectivement, dans le domaine de la statistique administrative, dans le domaine de la statistique appliquée aux sciences économiques et dans le domaine de la statistique appliquée aux sciences agronomiques, ont notamment conduit à des travaux importants et d'un haut niveau mathématique dans le cadre de la statistique universitaire. Il faut noter que ce sont ici les besoins de la pratique qui ont rendu la théorie plus féconde.

De même, on peut mettre en évidence d'importants transferts de méthodologies statistiques d'un *domaine d'application* à l'autre. Ainsi, la théorie des plans d'expérience a

très largement débordé le cadre agronomique ou biologique et est couramment utilisée dans de nombreux autres secteurs de la statistique appliquée : laboratoires, industries, etc. Ainsi également, l'analyse factorielle, d'origine essentiellement psychométrique, et la théorie des équations de régression simultanées, d'origine économétrique, comme je l'ai dit précédemment, ont été utilisées aussi, par exemple, dans le domaine de l'écologie et de l'étude des peuplements forestiers.

**

En ce qui concerne la distinction que j'ai faite entre *la statistique du riche et la statistique du pauvre*, il est certain aussi que les pays développés peuvent apporter une contribution importante aux pays en voie de développement, et cela même, peut-on dire, à relativement « peu de frais ». Si on s'en tient aux chiffres que j'ai cités pour la Société Internationale de Biométrie, on peut en effet constater facilement qu'il suffirait que chaque membre de cette société qui appartient aux pays développés consacre chaque année trois semaines seulement de son temps aux pays en voie de développement pour doubler le potentiel que constitue précisément l'ensemble des membres des pays en voie de développement. Et d'une façon plus générale, je suis tenté de dire qu'il suffirait chaque année de 10, 15 ou 20 jours de travail de chacun des statisticiens des pays développés pour doubler le potentiel statistique des pays en voie de développement.

Mais inversement, il est vrai également que les pays pauvres peuvent apporter beaucoup aux pays riches, ne serait-ce que par leurs nombreux problèmes spécifiques et originaux. Dans le domaine agronomique par exemple, on peut citer notamment les problèmes que posent la plus grande hétérogénéité des terres et du milieu, le nombre plus élevé des plantes cultivées et la plus grande diversité de leurs variétés, la diversité des modes de culture, tels que les cultures mélangées ou intercalaires, etc.

**

Enfin, si on considère les différentes écoles, telles que la statistique inférentielle classique et l'*analyse des données*, il faut bien admettre que des échanges fructueux pourraient être réalisés à ce niveau également et sont d'ailleurs, dans une large mesure sans doute, en voie d'être réalisés. Ainsi, l'idée d'introduire des individus supplémentaires à l'issue d'un traitement de données déjà effectué, idée qui me semble constituer un des principaux apports de l'analyse factorielle des correspondances, pourrait être étendue à l'analyse factorielle classique, à l'analyse des composantes, voire aussi à d'autres méthodes de l'analyse statistique multivariée.

Inversement, l'analyse des données, au sens de l'école française, pourrait être considérablement enrichie si on arrivait à éliminer, dans certains cas au moins, son caractère exclusivement descriptif et non inférentiel. Ce qui ne signifie nullement, dans mon esprit, qu'il faille enserrer l'analyse des données dans le carcan gaussien.

Et, avant de conclure, je voudrais souligner encore le fait que l'*ordinateur*, que nous utilisons presque tous quotidiennement, au moins dans les pays développés, et la *télématique*, qui ne fera qu'augmenter la demande de méthodes statistiques, constituent et constitueront de plus en plus un lien supplémentaire entre les différentes formes de statistique ou entre les différentes « races » de statisticiens, si je peux me permettre d'utiliser cette expression. Et cela, entre autres, par des échanges de programmes, de

sous-routines, etc., échanges qui se développent considérablement depuis un certain temps.



Ma *conclusion* sera donc que les statisticiens ont tout intérêt à exploiter au maximum la diversité de leur discipline, pour ceux qui ressentent le caractère enrichissant de cette diversité, ou à surmonter les différences qui existent entre eux, pour ceux qui croient se heurter à des barrières ou qui, parfois, dressent des barrières, consciemment ou inconsciemment.

Cette conclusion implique que chacun d'entre nous, de même que chacun de nos successeurs ou de nos élèves, soit formés à la fois en matière statistique et dans l'un ou l'autre secteur d'application, que des échanges soient organisés entre les différents types de statistique auxquels j'ai fait allusion, que chacun ait la volonté de regarder ce qui se passe chez le voisin et, aussi, de comprendre le voisin, et cela, comme je crois l'avoir bien fait sentir dans mon exposé, sans oublier les pays les plus pauvres.

Mais cette conclusion soulève aussi quelques problèmes plus généraux. D'une part, qui dit « comprendre le voisin » et bien sûr, corrélativement, se faire comprendre par le voisin, me semble dire aussi faire l'effort de comprendre la langue du voisin et faire l'effort de s'exprimer lui-même correctement dans sa propre langue, parlée et écrite. Ce qui n'est malheureusement pas toujours le cas, loin s'en faut, notamment parmi les plus jeunes.

D'autre part, et je m'adresse ici aux plus anciens, qui dit « échanges », sous forme de stages, de recyclages, de missions à l'étranger, de résolutions de problèmes concrets, etc., dit également perte de temps, au moins à court terme. Or, il nous faut bien reconnaître que nombre de nos jurys de recrutement, de sélection, de promotion n'encouragent guère de telles « pertes de temps ». Dans ce sens, la statistique relativement large et unie que je souhaiterais voir se développer nécessiterait sans doute une évolution non négligeable des idées, notamment dans le corps enseignant des universités et de certaines grandes écoles et au niveau des cadres dirigeants de certaines institutions de recherche.

Fondamentalement, je crois donc que, sans vouloir privilégier une statistique par rapport à une autre, et cela dans quelque sens que ce soit, nous devons nous efforcer de donner des chances égales aux différentes formes de statistique dont j'ai essayé de vous faire saisir et la diversité et l'unité, tout en favorisant au maximum les échanges entre les différents types de statistique.