

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

M. DUMAS

Le statisticien, cet artiste ou de quelques définitions du mot statistique

Journal de la société statistique de Paris, tome 96 (1955), p. 87-95

http://www.numdam.org/item?id=JSFS_1955__96__87_0

© Société de statistique de Paris, 1955, 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/>

V

LE STATISTICIEN, CET ARTISTE

OU

DE QUELQUES DÉFINITIONS DU MOT STATISTIQUE

I. — UNE DÉFINITION « OFFICIELLE »

Une définition du nom féminin « Statistique », — définition quasi officielle puisque, presque seule, elle figure dans toute la population des « Larousse » du plus petit jusqu'au plus gros — est la suivante :

Statistique : Science qui a pour objet le classement méthodique des faits sociaux qui se prêtent à une évaluation numérique (impôt, recrutement...).

Puis adjectivement : qui a rapport à la statistique.

Nous pouvons, dès l'abord, nous déclarer d'accord avec la définition de l'adjectif; mais que penser de cette définition du substantif?

Manifestement d'abord, le mot « sociaux » y est aujourd'hui de trop; le domaine de la statistique, sans doute initialement borné aux faits sociaux, est aujourd'hui beaucoup plus vaste; nous pouvons le constater sans insister.

Il s'agit, d'après la définition, de faits qui se prêtent à une évaluation numérique. Cela n'est-il pas trop restrictif? et cela comprend-il bien, sans ambiguïté, certains faits qui, individuellement, ne se prêtent à aucune évaluation numérique? Si l'on s'occupe des couleurs d'yeux, ce qui peut être classé méthodiquement ce sont les faits dont chacun est la couleur des yeux d'un individu; après classement, mais après seulement, apparaissent des évaluations numériques qui sont alors des dénombrements. Sur ce point encore je ne désire pas insister, tout en remarquant que l'on est plus sûr d'être complètement compris lorsque l'on est très explicite que lorsqu'on ne l'est pas.

Je suis choqué par l'expression « science qui a pour objet le classement méthodique... » : puisque le classement est l'action de classer, cela semble ravalier dans ce cas particulier la science à l'exécution d'une besogne matérielle; la science vaut mieux que cela.

Ainsi aucun point de la définition « officielle » ne trouverait grâce à mes

yeux ; c'est pour le moins peu modeste de ma part de prendre une telle position, malgré les bonnes raisons que je peux mettre en avant ; pour me faire pardonner je n'ai qu'une chose à faire, c'est à m'exposer à vos critiques ; je vais le faire largement.

II. — PREMIÈRES DÉFINITIONS PROPOSÉES

J'estime que le substantif *statistique* a plusieurs acceptions pour chacune desquelles je vais proposer une définition ; je vais distinguer ces acceptions les unes des autres en adjoignant un indice littéral au mot statistique ; sans plus tarder je vais proposer une première série de trois définitions.

Statistique A : Science des méthodes pouvant servir au classement ou à la présentation, numérique ou graphique, sous une forme éventuellement résumée, d'un ensemble de valeurs numériques.

Pour évoquer des méthodes que l'on doit à la statistique A, citons :

- répartition des valeurs en classes ;
- définition de valeurs typiques et d'indices de dispersion d'un ensemble de valeurs numériques ;
- histogrammes, graphiques à axes de coordonnées à échelle linéaire, logarithmique, etc...

Statistique B : Art d'appliquer les enseignements de la statistique A à des valeurs numériques correspondant respectivement :

- soit aux intensités d'un même caractère chez les individus d'une certaine population ;
- soit aux effectifs des individus qui, dans une certaine population, présentent un même caractère qualitatif.

Par exemple, la population en cause est une population d'objets matériels que l'on a cherché à produire identiques entre eux ; le caractère est la longueur séparant deux points matérialisés sur chacun des individus de la population ; les valeurs numériques sont celles de ces longueurs.

Par exemple encore la population est constituée par l'ensemble des habitants de la France au 1^{er} janvier 1955 ; le caractère est le sexe, les valeurs numériques sont au nombre de deux, celles des effectifs respectifs des hommes et des femmes.

Statistique C : Tableau numérique, ou graphique, auquel a conduit une opération de statistique B.

III. — COMMENTAIRES

Je crois que chacune des trois définitions proposées correspond à une acception dans laquelle le mot statistique est employé de nos jours : le théoricien fait de la statistique A quand il imagine de nouveaux indices susceptibles de résumer certains ensembles de faits, ou encore quand il imagine de nouvelles formes et échelles de graphiques susceptibles d'assurer dans de bonnes conditions la présentation de certains ensembles de valeurs numériques. C'est là un travail abstrait qui mérite bien d'être rattaché à une science.

Je ne crois pas avoir à parler de la statistique B si ce n'est pour préciser que « population » désigne un ensemble d'individus possédant le caractère que l'on considère et pour préciser également ce qui suit : d'après ces définitions,

l'homme qui recueille des données (valeurs numériques ou éléments des dénombrements futurs — même si à ces valeurs ou éléments on doit ultérieurement faire application des enseignements de la statistique A —) ne fait pas de la statistique : il est *expérimentateur* ; il est *enquêteur*, mais il ne fait pas œuvre de statisticien.

Une question analogue mérite d'être examinée de très près : celle qui concerne l'homme qui prépare une de ces expériences, une de ces enquêtes : cet homme fait-il de la statistique ? J'ai tendance à répondre « non » (si cela devait être « oui », une définition devrait être ajoutée aux trois que j'ai énoncées) car je crois que « l'homme qui prépare... » a davantage à être bien au fait de la statistique A (et d'ailleurs aussi d'une statistique D dont je me réserve de parler un peu plus loin) mais que la qualité essentielle qu'il doit manifester dans cette préparation, c'est sa connaissance technique du milieu qui subira l'expérience ou l'enquête : c'est là une qualité de technicien.

Quant à la statistique C, ce n'est plus ni une science ni un art : c'est un résultat. La statistique de la population française en fonction de l'âge ne peut être consultée dans des ouvrages spécialisés que parce que des savants ont élaboré des méthodes de présentation et que des artistes ont choisi pour la présenter une de ces méthodes de préférence à toute autre.

Une comparaison va montrer combien la langue française est pauvre pour ce qui touche à la statistique ; si elle était aussi pauvre en ce qui concerne l'arithmétique (science des nombres) c'est encore le mot arithmétique qui serait employé à la place de calcul (art de résoudre les problèmes de l'arithmétique) et c'est encore le mot arithmétique qui serait employé à la place de « résultat de calcul ».

IV. — STATISTIQUE MATHÉMATIQUE

Par les trois définitions qui précèdent, je pense avoir épuisé, et au delà, tout ce qui se trouve dans la définition quasi officielle à laquelle je faisais allusion plus haut. Mais je n'ai pas épuisé ce qu'il y a dans le substantif « statistique » puisque je n'ai pas abordé la partie de la statistique qui est en rapport avec cette branche des mathématiques qu'est le calcul des probabilités et qui correspond à peu près à ce que l'on nomme parfois « statistique mathématique ».

Je vous propose d'emblée une formule qui peut paraître assez révolutionnaire :

Statistique D : Art d'utiliser le calcul des probabilités à la préparation des décisions.

Mon propos est de chercher à justifier cette formule. Je vais pour cela exposer un exemple simple qui met en jeu les personnages que l'on rencontre à l'occasion, je l'estime, de tous les problèmes de statistique en rapport avec le calcul des probabilités. Voici cet exemple :

Le calcul des probabilités nous enseigne (la statistique n'a rien à voir là-dedans) que si la probabilité d'arrivée d'une blanche est de $1/2$, la probabilité de l'arrivée de plus de 30 blanches en 50 essais est de 0,0595, soit pratiquement 0,060. Un joueur désirerait savoir, avant de se décider à commencer une partie de pile ou face, quelle probabilité il a de voir arriver plus de 30 fois pile en

50 essais; il s'adresse pour cela à un statisticien. Ce dernier examine la question; s'il se décide à schématiser l'arrivée de pile dans le jeu de pile ou face par l'arrivée d'un événement de probabilité $1/2$, il utilise le résultat de calcul de probabilité rappelé ci-dessus et déclare au joueur qu'il évalue à 0,060 la probabilité que ce joueur demandait à connaître. Le joueur note cela à titre d'élément d'appréciation puis, joignant cet élément à ceux qu'il pourrait avoir par ailleurs il prend sa décision : commencer la partie ou ne pas la commencer.

On remarque que l'exemple fait intervenir trois hommes ayant chacun un rôle bien défini.

Il y a d'abord le *joueur*, à savoir celui qui a à prendre une décision et qui demande à recueillir des éléments d'appréciation avant de la prendre; certains de ces éléments n'ont rien à voir avec la statistique quel que soit l'indice littéral dont on affecte ce mot; certains de ces éléments peuvent être demandés à une statistique C : ni les uns ni les autres ne nous intéressent ici; enfin certains de ces éléments — et c'est ceci qui justifie le nom de « joueur » par lequel je désigne le demandeur — sont en rapport avec un certain caractère aléatoire : leur élaboration est proprement du ressort de la statistique D.

Le *statisticien* s'empare du problème qui lui est posé par le joueur. Il cherche un modèle mathématique convenant aux données et en retient un; par exemple il assimile à une certaine variable aléatoire un caractère de ces données, ou encore un résumé des données, puis il pose le problème; celui-ci peut d'ailleurs être du ressort des probabilités directes ou des probabilités inverses, peu importe en l'occurrence.

La parole est ensuite au *mathématicien* qui découvre au bout de calculs, qui sont du ressort exclusif du calcul des probabilités, la valeur constituant la solution du problème. Le joueur dûment averti à la fois des conditions dans lesquelles ont été choisies les données du problème que le statisticien a posé au mathématicien et de la réponse faite par ce dernier, est en possession de l'élément d'appréciation qu'il recherchait.

Certes les trois hommes peuvent être réunis en un seul; du moins, c'est très souvent le statisticien qui a la charge matérielle des calculs qui ressortissent en propre au mathématicien; mais il n'y a rien là qui soit de nature à faire changer la définition que j'ai proposée pour la statistique D et qui me paraît heureuse. Cette définition commence par le mot « art »; je crois que cela aussi est heureux; ce mot évoque, en effet, une idée de liberté qui convient bien au cas qui nous occupe. Le statisticien dans le rôle qui lui incombe d'utiliser le calcul des probabilités doit mettre à profit cette liberté pour retenir parmi tous les modèles probabilistes imaginables, celui qu'il juge correspondre au point de vue le plus propre à ne pas induire le joueur en erreur; éventuellement il recherche plusieurs points de vue à chacun desquels correspondra une solution; plus il aura de cordes à son arc pour aller au-devant des désirs du joueur en matière d'éléments d'appréciation et mieux il saura utiliser ces cordes, plus il sera grand « artiste ».

Certes, le statisticien ne doit pas aller jusqu'à utiliser ses cordes avec excès; si, comme un vrai artiste, il peut légitimement marquer ses réponses de son *chic* personnel, il doit éviter de les marquer de ses *tendances* personnelles, comme ce lui serait souvent aisé.

V. — L'ÉLABORATION DES RÈGLES PROBABILISTES

Arrivé à ce point, je serais assez tenté de dire que j'en ai fini avec les définitions du mot statistique. Mais j'ai scrupule à passer sous silence une activité sur laquelle je vais m'arrêter quelque peu : dans la définition de la statistique D j'ai fait la part du mathématicien; mais cette part est parfois pratiquement très réduite parce que certains théoriciens ont mis au point, à partir des enseignements du calcul des probabilités, des règles pouvant être suivies aisément chaque fois que le problème posé est d'un type bien déterminé. Le théoricien, metteur au point de méthodes, fait-il de la statistique? Si oui, il fait de la statistique ne méritant aucun des indices A, B, C ou D.

La réalité est que le théoricien résout un problème dont les données lui sont peut-être suggérées par des genres de décisions qu'il y a parfois à prendre, mais un problème qui est du ressort exclusif du calcul des probabilités. Toute la question est de savoir si ce théoricien est un statisticien ou un probabiliste; elle n'apparaît pas comme de première grandeur.

Pour ceux qui penchent pour le « statisticien », je propose :

éventuellement statistique E : Calcul des probabilités, dans la mesure où cette science se propose l'élaboration de méthodes pouvant être utilisées à la préparation des décisions.

Pour ma part, je penche pour le « probabiliste » parce que je ne vois pas la nécessité de faire subir au calcul des probabilités la dichotomie que marque la définition de la statistique E.

VI. — DIGRESSION SUR LE PASSÉ

Maintenant que je n'ai plus aucune définition à vous proposer (ce n'est pas que la question soit épuisée), j'aimerais me pencher un peu sur le passé. Je n'ai aucune compétence pour vous entretenir d'étymologie; j'ai seulement entendu parler d'un verbe grec « statizein » qui aurait voulu dire « constater ». Par contre j'ai bien présente à la mémoire cette séance du 20 octobre 1954 où, après un magnifique exposé de M. Penglaou, j'ai eu sans aucun doute le grand tort de faire dévier la question sur les diverses acceptions du mot statistique. Mon intervention a eu le mérite de faire resurgir quelques souvenirs, vieux de quinze ans, que je partage notamment avec M. Prot, mais elle était fort téméraire car, mis ce jour-là au pied du mur, j'aurais été bien en peine de préciser quoi que ce soit sur ce substantif statistique.

Je pense m'être acquitté de la dette morale contractée du fait de mon intervention téméraire en vous faisant part de l'état de mes réflexions sur ce sujet. Mais entre mon intervention d'octobre dernier et mes réflexions toutes récentes, il s'est produit un fait dont je tiens à faire état : j'ai pris connaissance dans une revue technique d'un article qui a été pour moi une sorte de révélation, parce qu'il exposait dans un français de tous les jours des idées simples que j'avais rencontrées bien des fois, mais que je n'avais jamais pu dégager de l'obscurité où les ensevelissaient des termes d'aspect savant ou d'aspect étranger. C'est ainsi que les « recherches opérationnelles » devenaient pour cet auteur « la préparation scientifique des décisions ». Tout s'éclairait à mes yeux;

je m'aperçus que j'avais fait longtemps des recherches opérationnelles sans le savoir et de fil en aiguille j'en suis arrivé à penser que la statistique, dans ses rapports avec le calcul des probabilités, n'était autre que des recherches opérationnelles : ainsi est née petit à petit la notion de statistique avec l'indice D : si je n'avais pas eu cette définition à soumettre à votre appréciation, je n'aurais certes pas demandé à prendre la parole.

J'avais une double raison pour vous raconter cette petite histoire. J'avais d'abord plaisir à saluer l'auteur M. René de Chantal, que vous connaissez bien, ne serait-ce que pour l'avoir admis depuis peu dans notre Société. J'avais également plaisir à souligner l'intérêt majeur qui s'attache à ce que nous tous, si nous désirons faire œuvre utile, c'est-à-dire si nous désirons nous comprendre entre nous, nous faire comprendre de ceux qui ne demandent qu'à être, eux aussi, des adeptes de notre discipline et nous faire comprendre de ceux qui ne demandent qu'à avoir recours à notre art, l'intérêt, disais-je qui s'attache à ce que nous nous exprimions clairement avec des mots simples. Nous devons nous défier notamment de ces mots étrangers qui ne sont jamais compris en France avec toutes les nuances qu'ils ont dans leur langue d'origine, ou de ces traductions qui sont trop littérales pour n'être pas trompeuses.

VII. — LA POURSUITE DE L'ÉTUDE DE DÉFINITIONS CONVENABLES

Pensons un moment à l'avenir. Notre Société serait tout à fait désignée pour mettre au point un texte précisant ce qu'il faut entendre à l'époque actuelle par le mot statistique qui figure dans son titre; la réunion d'aujourd'hui pourrait faire avancer la question si la discussion qui va s'engager est réellement fructueuse; quoi qu'il en soit, à supposer ce problème résolu, il n'y aurait plus qu'à répandre ce texte parmi les industriels et les professionnels intéressés; les industriels sont en petit nombre : ce sont les éditeurs de dictionnaires français; les professionnels ne sont pas plus de quarante, d'après le règlement de l'Académie.

Quant à nous, qui nous appliquons à faire dire aux nombres tout ce qu'il n'est pas déraisonnable de leur faire dire, nous continuerions comme par le passé avec, en plus, la satisfaction de nous dire que nous possédons enfin un critère nous permettant de reconnaître en quoi nous sommes statisticiens et en quoi nous ne le sommes pas; et si, finalement, vous reprenez quelque chose de ces définitions que je vous ai proposées aujourd'hui, en quoi nous sommes des « artistes ».

QUELQUES DÉFINITIONS DU MOT " STATISTIQUE "

RÉSUMÉ

a) DÉFINITION « OFFICIELLE »

Science qui a pour objet le classement méthodique des faits sociaux qui se prêtent à une évaluation numérique (impôts, recrutement...).

b) PREMIÈRES DÉFINITIONS PROPOSÉES

Statistique A :

Science des méthodes pouvant servir au classement ou à la présentation numérique ou graphique, sous une forme éventuellement résumée, d'un ensemble de valeurs numériques.

Statistique B :

Art d'appliquer les enseignements de la statistique A à des valeurs numériques correspondant respectivement :

- soit aux intensités d'un même caractère chez les individus d'une certaine population;
- soit aux effectifs des individus qui, dans une certaine population, présentent un même caractère qualitatif.

Statistique C :

Tableau numérique ou graphique auquel a conduit une opération de statistique B.

c) STATISTIQUE MATHÉMATIQUE

Statistique D :

Art d'utiliser le calcul des probabilités à la préparation des décisions.

Éventuellement : Statistique E :

Calcul des probabilités dans la mesure où cette science se propose l'élaboration de méthodes pouvant être utilisées à la préparation des décisions.

M. DUMAS,

*Ingénieur en chef de l'Artillerie navale R.
Délégué général du Syndicat M. T. P. S.*

DISCUSSION

M. PENGLAOU. — Après la lumineuse communication de M. Dumas, un point important est désormais acquis : la caractérisation des *données* ne ressortit pas au domaine spécifique de la statistique. C'est affaire du spécialiste, du savant.

Cette prise de position n'a pas qu'une valeur théorique. Adoptée sur le plan de l'action : description, limitation, recensement des données (toutes démarches opérationnelles que l'auteur de la communication englobe sous l'appellation d' « enquête »), elle écartera définitivement les équivoques, les sophismes, les erreurs qui entachent trop de travaux prétendument objectifs. Ces conclusions semblent acquises définitivement : ma communication n'a pas été vaine puisqu'elle a suscité le remarquable travail de M. Dumas.

Ceci posé, il reste à apprécier les définitions proposées. S'il est normal et usuel de constater qu'un vocable peut comporter plusieurs acceptions, il convient cependant de rechercher une formule synthétique qui puisse les résumer toutes. Les cinq définitions présentées, si elles ont le mérite de couvrir toutes les démarches du statisticien, comportent des éléments dont la conciliation paraît difficile.

La statistique est apparentée tout à la fois à la science et à l'art. Il faut opter. De *science*, il n'en peut être question; ce n'est pas affaire de statisticien, on vient de le dire. Le mot *art* est amphibologique quand il ne s'applique pas à l'une des activités socialement délimitées, telle que la peinture, la musique, la littérature, etc... Mais un mot résume à merveille l'action décrite par les définitions A et B, à savoir la *technique*, qui se réfère d'une part aux notions acquises en l'espèce par la science et aux procédés mis en œuvre par le statisticien pour remplir le but qu'il s'est proposé et où les qualités d'intuition, d'ingéniosité entrent en jeu, outre le savoir-faire du mathématicien.

Et voilà ce qui constitue de surcroît des difficultés pour moi. Et de poser

la question : est-ce que la mathématique doit être rejetée des définitions A, B et C? « ...Présentation numérique ou graphique... d'un ensemble de valeurs numériques » n'implique-t-elle pas la corrélation mathématique des valeurs présentées? Pourquoi réserver à la seule rubrique « Statistique mathématique » l'utilisation du calcul des probabilités?

M. Penglaou, après avoir de nouveau rendu hommage à la brillante communication de M. Dumas, estime qu'une synthèse doit être tentée avec les éléments amassés avec tant d'ingéniosité par son collègue qui aura beaucoup fait pour clarifier les conceptions statistiques.

M. André VÈNE rappelle que la statistique est essentiellement une science historique, une science d'investigation. Il a conservé le souvenir d'une ancienne définition de la statistique suivant laquelle elle est la science des faits susceptibles d'être comptés. Contrairement à l'opinion de M. Dumas, il ne croit pas que l'on puisse réserver le nom de statistique aux commentaires sur les faits ou à leur interprétation et en exclure le travail d'inventaire et de dénombrement. Autant vaudrait exclure d'une définition de la science les observations en laboratoire des chimistes et des physiciens.

Au moment de rédiger cette note, M. André Vène retrouve dans un article du Dictionnaire d'Économie politique d'Inglis Palgrave trois définitions anciennes. La première est de notre ancien Président Émile Levasseur : « La statistique est l'étude numérique des faits sociaux. » L'autre est de l'historien allemand Lexis : « La statistique est l'examen mesuré en chiffres de la vie des hommes en société. » La troisième est, d'après l'auteur de l'article I. S. Leadham, une définition courante : « La statistique est l'étude des phénomènes sociaux qui peuvent être comptés et mesurés en chiffres. »

M. Dumas a eu tout à fait raison d'attirer l'attention de nos collègues sur l'insuffisance des définitions qui bornent la statistique aux faits sociaux. Cette limitation excessive a une explication qui est d'ordre historique. Apparue dès l'antiquité comme un élément d'information nécessaire aux gouvernements, la statistique a été longtemps utilisée, presque exclusivement, par les économistes, les spécialistes des sciences politiques et les dirigeants des administrations. Elle a été, depuis lors, étendue à d'autres objets.

Par une large extension, l'on arriverait à la définition suivante, provisoire peut-être, mais acceptable : « La statistique est le dénombrement et l'étude chiffrée des faits sociaux, biologiques et naturels dont les proportions et les variations sont susceptibles de mesure. »

M. DUMAS répond qu'il a cherché à faire un travail d'analyse dans le but essentiel de bien distinguer les unes des autres les différentes acceptions du mot statistique; la chose n'était pas inutile puisque dès l'abord la discussion s'est engagée, avec les deux distingués intervenants, relativement à un point important, celui du recensement des données. Il faut que, préalablement, nous nous mettions d'accord sur ce point et sur quelques autres, pour que nous soyons en état de tenter de synthétiser les acceptions en un petit nombre de formules.

Sans doute, pouvons-nous dès maintenant discuter le point de savoir si science et art pourraient être correctement remplacés par le mot de technique; mais nous n'en sommes pas encore selon moi, à chercher comment inclure

dans une même définition à la fois ce qui fait intervenir le calcul des probabilités et ce qui ne le fait pas; c'est dire en particulier que la définition proposée finalement par M. Vène me paraît prématurée.

Par contre, M. Vène nous signale quelques définitions « anciennes »; cela va au-devant des désirs qu'a bien voulu m'exprimer un de nos anciens Présidents, M. Maurice Fréchet : intérêt qu'il y aurait à recueillir les différentes définitions de la statistique qui ont été données par les nombreux auteurs sur ce sujet; une liste analogue a été dressée pour le mot économétrie et publiée par Tintner.
