



SCIENCES ÉCONOMIQUES

ANDRÉ PIATIER

STATISTIQUE

1

PRESSES UNIVERSITAIRES DE FRANCE

THÈMES

SCIENCE ÉCONOMIQUE
SCIENCE ÉCONOMIQUE

ANDRÉ PIATIER

Docteur en Sciences Économiques

Professeur à l'École Supérieure de Commerce de Paris

Chargé de Cours à l'École des Hautes Études Commerciales

Chargé de Cours à l'École des Hautes Études Commerciales

STATISTIQUE

1

23523

160R
13346
(H)

STATISTIQUE

I

1873

1873
1874
(A)

THÉMIS

COLLECTION DIRIGÉE PAR MAURICE DUVERGER
SCIENCES ÉCONOMIQUES

ANDRÉ PIATIER

*Agrégé des Facultés de Droit
et des Sciences économiques
Professeur à l'École Pratique des Hautes Études
Directeur du Centre d'Étude des Techniques
économiques modernes*

Statistique

1 / Statistique descriptive et
Initiation à l'analyse



PRESSES UNIVERSITAIRES DE FRANCE

108, BOULEVARD SAINT-GERMAIN, PARIS

DL -- 1 4 1970.02903

THÉMIS

COLLECTION ORGANISME PAR MAURICE BASTIENNE
SCIENCES ÉCONOMIQUES

ANDRÉ PIATIER

Professeur de l'Université de Paris
Directeur de l'Institut de Recherches Économiques
de l'Université de Paris

DU MÊME AUTEUR

Statistique et observation économique (Statistique, Économétrie, Conjoncture, Comptabilité nationale), 2 vol., Presses Universitaires de France, 1961.

L'équilibre entre le développement économique et le développement social, 1 vol., Paris, Médicis, 1963.

Structure et perspectives de la consommation européenne, 1 vol., Paris, Sélection, 1968.

En collaboration avec P. CAHUZAC et L. CHAMBADAL, dans la collection « Thémis » :

Economie et mathématiques (Éléments et exercices) :

T. 1 : *Les nombres. Algèbre. Analyse*, Paris, Presses Universitaires de France, 1965.

T. 2 : *Analyse statistique et applications à l'économie*, Paris, Presses Universitaires de France, 1970.



Dépôt légal. — 1^{re} édition : 3^e trimestre 1962

3^e édition mise à jour : 1^{er} trimestre 1970

Tous droits de traduction, de reproduction et d'adaptation réservés pour tous pays

© 1962, Presses Universitaires de France

INTRODUCTION

Le programme officiel

Le programme du cours de première année de Statistique (Facultés de Droit et des Sciences Economiques) a été fixé par arrêté du 19 août 1960 (*J.O.* du 25 août 1960).

Il comporte les rubriques suivantes :

I. *L'élaboration des statistiques :*

- La statistique et son domaine d'application ;
- Les problèmes posés par l'observation des faits économiques ;
- Les différentes formes de la connaissance statistique ;
- Les organismes de la statistique ;
- Les sources et les publications.

II. *La présentation des statistiques :*

- Signification et validité des données chiffrées
- Erreurs et approximations ;
- Tableaux statistiques ;
- Graphiques et diagrammes.

III. *Analyse statistique élémentaire :*

- Les moyennes ;
- Les écarts ;
- Les nombres indices ;
- Notion de corrélation simple ;
- Introduction à l'analyse des séries chronologiques ;
- La portée du calcul des probabilités en matière de sondage et d'estimation.

Contenu du Manuel

Nous nous sommes efforcés de suivre d'aussi près que possible le programme officiel. Si le découpage en titres et en chapitres est un peu différent de celui qui est proposé, l'ensemble des questions mises au programme est traité.

Pour tout enseignement nouveau, il importe de savoir avec quelle précision et avec quel appareil technique les matières doivent être traitées. Nous avons tenu compte du fait que ce Manuel s'adresse à des étudiants qui sont tout au début de leurs études d'enseignement supérieur — et aussi du fait que ces étudiants sont loin d'avoir tous un bagage mathématique convenable. Certes, en première année un cours de mathématiques pour économistes leur est enseigné à côté du cours de statistique. Mais comme des décalages dans le temps peuvent se produire entre les deux enseignements — et comme les leçons de mathématiques peuvent demander un certain délai d'assimilation —, nous avons estimé que ce premier cours de statistique ne devait, en règle générale, faire appel qu'à des connaissances mathématiques limitées, largement acquises au niveau du baccalauréat.

Le but de la nouvelle licence ès sciences économiques est d'attirer vers l'étude scientifique de l'économie des jeunes de formation différente et de leur donner les moyens de mener à bien cette étude. Le programme officiel tient compte de cette double nécessité — de ne pas rebuter les débutants et de graduer les efforts, puisque, pour la première année, il prévoit deux parties bien distinctes : Statistique descriptive d'une part, Introduction à l'analyse statistique d'autre part.

Le Manuel est consacré à l'étude *des statistiques*, c'est-à-dire des documents de base obtenus après mesure des phénomènes, réunion et assemblage des données. Les méthodes d'élaboration d'une *image du réel* qui constituent la *statistique descriptive* doivent être connues avant même que soit abordée la *statistique théorique* qui, avec des moyens mathématiques sans cesse accrus, s'efforce de découvrir les propriétés des ensembles et d'exploiter systématiquement ces propriétés.

Il ne faut pas oublier, comme le dit E. Morice, que « si la

statistique théorique fournit des moyens d'analyse d'une documentation chiffrée, les problèmes se posent relativement à des réalités concrètes plus ou moins difficiles à observer sinon à définir, et que les plus savantes analyses théoriques ne valent rien — ou même sont plus dangereuses qu'utiles — si la documentation chiffrée utilisée n'est pas une description suffisamment fidèle de ces réalités complexes ».

Le Manuel de première année est donc, pour l'essentiel, tourné vers la *description statistique* et il ne contient qu'une rapide introduction aux méthodes d'analyse, tout juste suffisante pour illustrer l'emploi qui peut être fait des données dont on aura étudié la collecte, le dépouillement et la présentation — et pour préparer les esprits au programme de deuxième année, entièrement consacré à la statistique probabiliste.

Ce Manuel restera donc un instrument de travail pour toute la durée des études — au moins pour les matières dont il ne sera certainement plus question dans les programmes des années suivantes — par exemple pour les méthodes d'élaboration statistique, recensements, construction d'un questionnaire et conduite des enquêtes, graphiques et cartogrammes, sources statistiques, mécanographie, indices.

BIBLIOGRAPHIE GÉNÉRALE

Elle est complétée par des bibliographies spéciales données à la fin des chapitres.

1) Ouvrages n'impliquant pas de connaissances mathématiques :

- A. VESSERAU, *La statistique*, « Que sais-je ? », n° 281, Paris.
- A. LIORZOU, *Initiation pratique à la statistique*, Paris, 1956.
- R. DUMAS, *L'entreprise et la statistique* (2 vol.), Paris, 1959.
- A. MONJALLON, *Introduction à la méthode statistique*, 1958.
- D. HUFF, *How to lie with Statistics*, New York.
- H. GUITTON, *Statistique et économétrie*, Paris, 1958.
- A. PIATIER, *Statistique et observation économique*, Paris, 1961.

2) Ouvrages demandant des connaissances élémentaires :

- E. MORICE et F. CHARTIER, *Méthode de statistique* (2 vol.) : I. *Elaboration des statistiques* ; II. *Analyse statistique*, I.N.S.E.E., 1954.
- R. DAVAL et G. Th. GUILBAUD, *La méthode statistique* (2 vol.) : I. *Statistique descriptive* ; II. *Distribution binomiale*, Paris, 1950-1951.
- F. C. MILLS, *Statistical methods applied to economics and business*, New York, 1924 et 1938.
- CRUM, PATTON et TEBBUT, *Introduction to Economics Statistics*, New York et London, 1939.
- YULE et KENDALL, *An Introduction to the theory of Statistics*, London, 1950.
- R. G. D. ALLEN, *Statistics for Economists*, London, 1949.
- R. A. FISCHER, *Statistical methods for research workers*, London, 1951 ; trad. franç. : *Méthode statistique pour la recherche*, Paris.
- D. A. S. FRASER, *Statistics : An Introduction*, New York, 1958.

3) Ouvrages demandant de sérieuses connaissances mathématiques :

- J. DESABIE, *Théorie et pratique des sondages*, Paris, 1966.
- E. MALINVAUD, *Méthodes statistiques de l'économétrie*, Paris, 1964.
- G. TINTNER, *Mathématiques et statistiques pour les économistes*, t. I, Paris, 1961.
- J. MOTHES, *Prévisions et décisions statistiques dans l'entreprise*, Paris, 1962.
- G. CALOT, *Cours de calcul des probabilités*, Paris, 1967.
- C. FOURGEAUD et A. FUCHS, *Statistique*, Paris, 1967.
- A. PASQUIER, *Éléments de calcul des probabilités et théorie des sondages*.
- A. PIATIER, P. CAHUZAC et L. CHAMBADAL, *Economie et mathématiques*.
T. 1 : *Les nombres. Algèbre. Analyse*, « Thémis », Presses Universitaires de France, Paris, 1965 ; t. 2 : *Analyse statistique et applications à l'économie*, « Thémis », Presses Universitaires de France, Paris, 1970.
- R. SCHLAIFER, *Probability and Statistics for Business Decisions*, New York, 1959.

TITRE PREMIER

VUE D'ENSEMBLE SUR LA STATISTIQUE

Toute étude scientifique suppose, au point de départ, une double démarche : d'un côté la *méthode axiomatique*, qui consiste dans l'emploi de déductions logiques tirées d'axiomes *a priori* sans faire appel à l'observation et à l'expérimentation, donne des *schémas modèles*, ou une *représentation abstraite* du monde dont les mécanismes sont « expliqués » par le raisonnement pur. De l'autre côté, la *méthode inductive* part de l'observation du monde sensible, ordonne et sélectionne les informations, les simplifie, les réduit, les confronte — et, d'effort en effort, arrive à des *schémas concrets* dont on découvre les normes et les régularités. Mais la véritable investigation ne se termine que par la rencontre des deux approches, que par la confrontation de leurs résultats et par l'enrichissement de l'une par l'autre.

Autrefois, chez les économistes, on opposait l'école théorique et l'école positive. L'apparition de l'*économétrie* a montré que chacune de ces deux écoles ne faisait qu'une partie du chemin — et qu'il fallait encore faire la synthèse entre les *lois abstraites* imaginées dans un travail de pure logique et les *régularités statistiques* dégagées par un patient travail d'analyse et de mesure. L'économétrie, conçue comme une conjugaison de la théorie économique et des mesures statistiques effectives, n'est que l'application dans une des sciences sociales d'une procédure utilisée dans les autres branches de la connaissance.

Nous venons de situer la statistique dans l'étude économique. Il s'agit maintenant d'envisager la statistique pour

elle-même : là encore nous retrouvons la dualité des approches et leur rencontre. L'étude des fonctions, l'analyse combinatoire, le calcul des probabilités aboutissent à des constructions de l'esprit, à des schémas dont il est bon de rapprocher les données fournies par l'observation de la réalité. Des deux branches de la statistique — statistique théorique et statistique descriptive du réel —, seule la seconde est étudiée systématiquement dans ce Manuel. L'histoire de la statistique nous montrera (chap. I^{er}) que cette branche est la plus ancienne. La première, née seulement au XIX^e siècle, s'est développée rapidement au XX^e et n'a pas tardé à « prendre contact avec le monde sensible... prouvant au-delà même de toute espérance que la connaissance est une » (Fréchet). Cette descente du monde des idées pures et des schémas logiques vers le réel pourra être envisagée dans les Manuels de deuxième et de troisième années. Dans le présent Manuel, il ne pourra guère être question que de la démarche symétrique : s'élever de l'observation banale et de la mesure concrète à des présentations schématiques de grandeurs épurées, simplifiées en vue de préparer la rencontre avec celles que propose la théorie statistique mathématique.

CHAPITRE PREMIER

HISTORIQUE ET DÉFINITION DE LA STATISTIQUE

La statistique s'est complètement transformée par l'emploi du calcul des probabilités. Un historique (section I) montre sa lente évolution jusqu'au XVIII^e siècle et sa mutation brusque lorsque les travaux de Pascal, Fermat, Moivre, Bernoulli, Gauss, Laplace, etc., ont pu orienter des réalisations nouvelles. La section II peut alors éliminer les mauvaises définitions de la statistique et tenter d'en donner une qui corresponde à ses moyens actuels.

SECTION I. — Les quatre phases de l'histoire de la statistique

Les préoccupations statistiques remontent à la plus haute Antiquité. Mais leur contenu a sensiblement varié avec le temps. Il est possible de distinguer quatre phases dans l'évolution historique.

§ 1. PREMIÈRE PHASE : LE DÉNOMBREMENT

Dès qu'une autorité politique se constitue, l'idée d'inventaire plus ou moins régulier de la population et des richesses existantes sur le territoire est liée à la prise de conscience de la souveraineté et aux premiers efforts d'administration.

Entre le IV^e et le II^e millénaire avant J.-C., les *Chinois* disposaient de renseignements sur la population et utilisaient des tables de statistique agricole. L'information du « chef » se rattache ici à la pratique de gouvernement et aux rites : selon l'expression de P. Mus, ces civilisations

s'appuyaient sur le calendrier — et les opérations de production, effectuées sous une incitation religieuse, ne commençaient qu'après les cérémonies et l'accomplissement des gestes rituels des autorités.

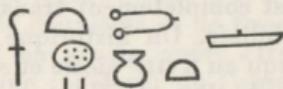
Les *Egyptiens* connaissaient aussi les recensements réguliers. La lecture des hiéroglyphes a même révélé l'influence considérable de cette « statistique administrative » sur la vie du pays et même sur sa langue.

Les années, en effet, n'avaient pas une expression numérique analogue à celle que nous connaissons aujourd'hui. Elles étaient définies, dès l'époque des premières dynasties, par les événements importants qui les avaient marquées — par exemple l'année où le pharaon a vaincu les Asiatiques, ou l'année de l'inauguration de tel temple à tel dieu.

L'expression « année » était représentée par :



ce signe étant complété par l'indication de l'événement dominant. Bientôt, on marque dans les annales l'année où s'était effectué le recensement pour l'impôt — et on dit



soit commencement du temps du 2^e recensement du bétail (ou des champs — ou de l'or). Au début, cette opération fiscale avait lieu tous les deux ans. Mais, comme à partir de la VI^e dynastie le recensement s'effectuait tous les ans, le mot *recensement* devint synonyme d'*année* (d'après Jean Capart, *Hiéroglyphes*, 1946).

Signalons encore que les *Egyptiens* avaient mis au point le plus ancien « baromètre économique » connu à ce jour — et qui était le *Nilomètre* : la hauteur de la crue du Nil était un excellent indice de fertilité — et elle servait à fixer le montant des impôts.

Les civilisations de la même époque (Ur et Indus) disposaient probablement de moyens statistiques analogues que les travaux archéologiques actuels permettront peut-être de mieux saisir.

La Bible est remplie d'indications statistiques et le IV^e Livre de Moïse donne un bon exemple de dénombrement des Israélites en état de porter les armes.

On y trouve même une appréciation curieuse sur le caractère « satanique » de la statistique (Chronique, 1-21 : « Mais Satan s'éleva contre Israël et incita David à faire le dénombrement. ») Malgré les protestations de Joab et les refus de répondre des tribus de Levi et de Benjamin, l'opération eut lieu et permit de dénombrer « d'Israël onze cent mille hommes qui tiraient l'épée et de Juda quatre cent soixante dix mille ». Mais « cette affaire déplut à Dieu » et, bien que David ait reconnu « son très grand péché », la punition divine fut trois jours de peste.

Les Grecs et les Romains multiplièrent les investigations analogues — les buts poursuivis n'étaient plus exclusivement fiscaux ou militaires.

Il s'agissait aussi de répartir les terres, de fixer la condition sociale des habitants, de faciliter le ravitaillement (par exemple stockage du blé par Pompée, pour nourrir 486 000 personnes) — distribution de secours (Auguste donnant en 5 avant J.-C. 60 deniers à chacun des 320 000 citoyens de la plèbe).

Mais les méthodes utilisées pour ces dénombrements semblent avoir été très changeantes ; elles dépendaient d'un objectif momentané — d'où les incertitudes sur la signification des résultats obtenus. C'est ainsi que J. Carcopino (*La vie quotidienne à Rome*, Hachette) montre qu'on peut hésiter pour l'évaluation de la population de Rome entre 250 000 et plus d'un million d'habitants, selon les sources consultées. Le fait que très peu de résultats statistiques de l'Antiquité nous soient parvenus s'explique moins par le dédain des historiens des siècles passés pour ce genre de documentation que par la destruction probable des données, une fois finie l'action administrative effectuée grâce à elles. Et pourtant il y eut dans la ville de Rome 69 recensements (ou *census* — opération faite en présence du censeur). À l'incertitude sur la définition des « unités » recensées, par exemple population totale y compris les esclaves ou seulement « citoyens », s'ajoutaient les incertitudes sur leur relevé : pour certains recensements, la population était obligée de se déplacer en masse pour rejoindre les lieux de naissance ; c'est à cause d'un recensement que le Christ est né dans une étable.

On sait que, dans la ville de Rome, plusieurs dénombrements se sont faits par immeubles et sous la responsabilité du propriétaire (par exemple en 44 avant J.-C. sous César). C'est la méthode encore employée le plus souvent aujourd'hui.

Selon Tacite, l'empereur Auguste aurait été le premier à faire un bilan des richesses de l'Empire romain (soldats, navires, ressources privées et revenus publics). Cette première ébauche de comptabilité, restée sans lendemain, fut présentée au Sénat par son successeur.

Tous ces exemples sont relatifs à des statistiques *d'état* (inventaire à un instant donné) ; les statistiques de *mouvement* (retracant une évolution) semblent n'avoir eu qu'une place très secondaire ; on sait cependant qu'on enregistrait dans certains temples naissance, puberté et mort des individus. C'est là l'ancêtre des registres des paroisses qui ont eux-mêmes précédé l'*état civil* créé au moment de la Révolution française.

Au Moyen Age, les pratiques statistiques se poursuivent et on cite par exemple les relevés nombreux faits à l'époque de Charlemagne et le *Domesday Book* de Guillaume le Conquérant. Mais comme le dit A. Julin (*Principes de statistique théorique et appliquée*, Bruxelles, 1921) : « Les matières sur lesquelles portèrent ces investigations font partie, incontestablement,

du domaine assigné à la statistique, mais il faut renoncer à trouver, dans les recherches entreprises dès l'Antiquité et le Moyen Age, aucun principe de technique et aucune idée de la nature scientifique de la statistique. »

Il est possible que de nouvelles recherches historiques montrent cependant que la statistique ancienne pouvait prétendre à plus qu'à des collectes de faits fragmentaires : les travaux de M. Louis Baudin sur les *Incas* nous découvrent des descriptions fort bien articulées et des calculs en vue d'un effort insoupçonné de planification (*Les Incas du Pérou, Médecis ; La vie quotidienne au temps des derniers Incas*, Hachette).

Mais même si la statistique atteignait ce stade dès la plus haute Antiquité, on doit reconnaître avec M. Vessereau qu'elle a conservé jusqu'au XVIII^e siècle un caractère passif : « La méthode d'interprétation fait encore défaut. »

§ 2. DEUXIÈME PHASE : DE LA DESCRIPTION DES ENSEMBLES A L'ARITHMÉTIQUE POLITIQUE

C'est à un Italien, F. Sansovino, qu'on attribue le mérite d'avoir, dans son livre de 1567 (*Du gouvernement et de l'administration*), envisagé la description chiffrée des Etats, de leurs richesses, de leur population, de leur commerce, de leur industrie et de leurs finances. La fin du XVI^e siècle est riche de travaux de ce genre. Jean Bodin dans la *République* prône l'emploi des renseignements numériques et l'organisation de dénombremens (1576). Sully illustre ses *Economies royales* de nombreux éléments chiffrés. P. d'Avity publie *Les Etats, Empires et principautés du monde*. En Allemagne, S. Münster fait un travail analogue qu'il appelle *Cosmographie*.

Les idées mercantilistes entraînent une intensification de ces recherches. Colbert multiplie les enquêtes sur les manufactures, le commerce et la population ; les intendants envoient à Paris leurs Mémoires. Vauban, plus connu par ses fortifications ou sa *Dîme royale* (première proposition d'un impôt sur le revenu), se signale comme le véritable précurseur des sondages. Un peu plus tard, Buffon se préoccupe de ces problèmes avant de se tourner vers l'histoire naturelle. Et le XVIII^e siècle se termine avec le plan de *Mathématiques sociales* de Condorcet.

Mais ce bouillonnement d'idées et de travaux en France aux XVII^e et XVIII^e siècles retient moins l'attention (parce que assez dispersé) que l'effort des écoles allemande et anglaise.

L'école allemande, créée par Conring (1606-1681), ne se signalerait pas spécialement — car elle ne sort pas de cette description des « ressources » et des « forces » des Etats, elle ne crée aucun élément chiffré nouveau et se contente de compiler, sans critique des sources, les documents existants — si l'un de ses représentants, G. Achenwall (1719-1772), professeur à Göttingen, n'avait forgé le mot *Statistik*. Peu importe que les commentateurs d'aujourd'hui en voient l'origine dans le mot latin *Status*, ou dans l'italien *Statista* : de toute façon la référence à la notion de *Politique* est évidente. Von Schlözer (1735-1809), de Göttingen également, poursuit le même travail beaucoup plus orienté vers l'information que vers la recherche des lois. Et en 1758, Büssching fait plus des rapprochements entre Etats que des comparaisons statistiques véritables.

L'école anglaise amorce un nouveau progrès en dépassant le stade purement descriptif. « On y analyse les données de l'observation, mettant ainsi en évidence certaines permanences. Se basant sur ces permanences et interprétant les données, on cherche des lois et on essaie même de faire de la prévision » (Morice). Ses trois principaux représentants sont Graunt (1620-1674), Petty, auteur de la fameuse *Political Arithmetic* (1623-1687) et Halley (1662-1742), astronome.

Tous vont calculer des proportions entre phénomènes, les classer, essayer de dégager des régularités. Comme la documentation en matière économique est fragmentaire et souvent contestable, leurs résultats dans ce domaine resteront très approximatifs. Au contraire, les travaux qu'ils feront sur des données démographiques seront plus valables. Utilisant les registres des décès de Londres, Graunt établit les premiers taux de mortalité par tranches d'âge ; il découvre ainsi la proportion plus grande de naissances masculines (107 contre 100 naissances féminines). Halley, en se basant sur cinq ans d'état civil de la ville de Breslau, préfigure les travaux d'actuariat moderne. Sa table de mortalité annonce la possibilité de transactions économiques nouvelles : assurances-vie, et contrats en viager.

Ces travaux ne restaient pas sans écho sur le continent. En France, Deparcieux, de la Société royale des Sciences de Montpellier, publie en 1746 un ouvrage trop peu connu : *L'essai sur les probabilités de la durée de la vie*

humaine. On lui doit les premières tables françaises de mortalité établies d'après les tontines fondées en 1689 et 1696 et les registres de décès de couvents parisiens dépouillés sur une longue période (1607-1745).

Avec lui, la critique préalable des données s'affine et devient systématique. En voici un exemple : surpris de constater dans les registres de Londres un excédent de décès sur les naissances et dans ceux de Paris un excédent de naissances alors que les conditions démographiques paraissaient semblables, il explique :

« A Londres, les mères nourrissent leurs enfants elles-mêmes et, par là, on a en général la mortalité de ceux qui y naissent et de ceux qui viennent s'y établir ; au lieu qu'à Paris, les mères ne nourrissent pas leurs enfants, on n'a pas la mortalité de ceux qui meurent en nourrice ; et le nombre en est grand. »

Avant d'être énoncées, les « régularités » sont recoupées au moyen de calculs faits sur d'autres sources (l'état des baptêmes et des morts de la paroisse de Saint-Sulpice de Paris pendant trente ans, par exemple). Et Deparcieux tire des conclusions confirmées depuis lors : « Les femmes vivent plus longtemps que les hommes... » « Plus de femmes veuves que d'hommes veufs » (non seulement à cause de la différence de longévité mais aussi parce que les hommes se remarient plus souvent). « On marie bien plus de filles avant l'âge de vingt ans que de garçons. » « On vit plus longtemps dans l'état du mariage que dans le célibat. » Le rapport des baptêmes de garçons à ceux des filles est « à très peu de choses près comme 24 est à 23 » (on dirait aujourd'hui le taux de masculinité), etc.

En Allemagne, le pasteur J. Süssmilch, en lutte avec l'école descriptive de Göttingen, fait à la même époque (1741) des constatations analogues. Mais il éprouve le besoin d'une référence divine : s'il naît 21 garçons pour 20 filles, l'égalité est établie à l'âge du mariage et Dieu assure ainsi la perpétuation de l'espèce humaine. En Suède, Wargentin fonde le même courant de recherches.

La fin du XVIII^e siècle est une période de démarrage des bureaux et organismes statistiques. Chaptal, ministre de l'Intérieur, lance, en 1801, le premier recensement général de la population, développe les études industrielles — partout l'évaluation des surfaces cultivées, des productions, des échanges devient systématique et, pendant les deux tiers de son cours, le XIX^e siècle connaîtra des investigations statistiques sans précédent. Une décadence de l'observation statistique s'amorcera, surtout en France, vers la fin du Second Empire : moindre intérêt de la part des hommes d'Etat, formulation plus intransigeante du libéralisme économique, disparition du contact avec les économistes plus soucieux de théorie et de micro-économie marginaliste, réaction contre un scientisme un peu imprudent ?

Il est difficile de préciser les causes de ce recul. Mais au début de la III^e République, la Statistique générale de la France n'est plus qu'un petit bureau occupé par six personnes...

§ 3. TROISIÈME PHASE : STATISTIQUE ET CALCUL DES PROBABILITÉS

Pendant que la description statistique des faits passait des savants à l'administration, une véritable révolution scientifique allait donner à la statistique son nouveau visage : parti de l'étude des jeux de hasard, le calcul des probabilités s'avère rapidement comme un instrument d'analyse extrêmement puissant pour l'étude des phénomènes économiques et sociaux — et d'une façon plus générale pour l'étude des phénomènes « à causes trop complexes pour qu'on pût les connaître toutes et les analyser » (R. Dumas, *La statistique et l'entreprise*).

Les étapes qui ont permis cette révolution avaient été : 1713 : publication d'*Ars Conjectandi* de J. Bernoulli qui énonce la Loi des Grands Nombres ; 1763 : publication du Mémoire de Bayes ; et 1812 : publication de la *Théorie analytique des probabilités* de Laplace. A la même époque, Gauss formule la loi normale et établit la fameuse courbe qui porte son nom.

C'est à Quételet (1796-1894) que revient le mérite de la transposition des nouvelles méthodes à l'économie et à la démographie : son œuvre décisive s'intitulait *Lettres sur la théorie des probabilités appliquées aux sciences morales et politiques*. Secrétaire perpétuel de l'Académie royale de Belgique, il se livre aussi à l'application des procédés statistiques à l'astronomie, à la météorologie et à la botanique. Mais sa réussite est surtout marquée dans « l'*Etude des qualités physiques, intellectuelles et morales des êtres humains* — en vue de constituer une sorte de physique sociale — d'après laquelle ces diverses qualités, considérées dans la masse, s'équilibreraient autour d'un être fictif, l'*homme moyen* » (Dugé de Bernonville, *Initiation*). C'est donc à lui qu'on doit le fameux *homo statisticus* qui était venu relayer l'*homo œconomicus* des classiques. Les humains « réels » se distribueraient autour de cet homme aussi abstrait que son prédécesseur, comme « au hasard » — ou comme les

erreurs dans la mesure d'une grandeur quelconque. Par cette découverte d'une répartition probabiliste des caractères physiques ou moraux des humains, Quételet ouvre la voie à la recherche de « lois » dans le domaine des sciences humaines.

Ses continuateurs de la fin du XIX^e siècle sont moins connus que lui. Il faut cependant retenir Lexis, del Vecchio et Rumelin. À partir du début du XX^e siècle, la statistique, qui avait failli n'être qu'une approche renouvelée des problèmes économiques et démographiques, va connaître une prodigieuse expansion dans la plupart des branches de la connaissance (voir chap. suiv.). Et, considérée comme une discipline *autonome*, au carrefour de toutes les sciences, elle s'enrichit de nouveaux apports.

§ 4. QUATRIÈME PHASE : STATISTIQUE ET CALCUL ÉLECTRONIQUE

L'irruption des ordinateurs dans l'activité moderne amplifie les possibilités de traitement des données dans deux directions : d'une part, des populations statistiques qu'il n'était pas question d'analyser autrefois du fait de leur importance, sont aujourd'hui facilement prises en compte. D'autre part, des procédés de calcul impraticables manuellement sont maintenant utilisés de façon courante. Enfin la statistique, grâce à ces possibilités nouvelles, utilise de nouveaux outils mathématiques (algèbre moderne, calcul matriciel, analyse combinatoire) et, dans le domaine des sciences économiques et sociales, elle apparaît de plus en plus soudée aux étapes ultérieures de toute étude : statistique, économétrie, conjoncture, recherche opérationnelle représentent souvent un ensemble indissociable.

SECTION II. — Définition de la statistique

Les paragraphes qui précèdent permettent de comprendre d'une part le caractère dérisoire de certaines définitions de la statistique — et d'autre part de les expliquer par la connaissance plus ou moins poussée que leurs auteurs avaient de cette discipline.

Ces définitions sont innombrables : il y en avait déjà 63 en 1870 et dans son article de 1935 (*Revue de l'Institut international de Statistique*) Willcox en dénombrait plus d'une centaine. L'inflation a continué depuis car « parmi les thèmes à propos desquels les statisticiens ne sont pas d'accord se trouve la définition de leur science » (Kendall). Une mise en ordre s'impose d'autant plus que bon nombre de manuels livrent ces définitions dès leurs premières pages et les imposent donc à un lecteur encore non informé. Un travail critique de classification permet de faire les distinctions suivantes :

§ 1. DÉFINITIONS HUMORISTIQUES ET CRITIQUES

Ce sont celles émanant de personnes n'ayant aucune connaissance de la méthode, de ses buts et de ses succès. Elles n'ont de résonance que chez ceux qui partagent cette ignorance. Elles ont donc une excellente valeur de « test ». La plus célèbre est celle qu'on attribue à Disraeli ou à Bismarck : « Il y a trois formes de mensonge, le mensonge ordinaire, le damné mensonge et la statistique. » A. Sauvy a fait une critique assez véhémement de cette définition et montré que le langage, la photographie et des quantités d'autres modes d'expression humaine pouvaient être tout aussi mensongers, sinon plus, que la statistique (Cours à l'Institut d'Etudes politiques, 1946-47). Macauley disait : « Les chiffres disent toujours ce que souhaite l'homme habile qui sait en jouer », et Thiers pensait que « la statistique est l'art de préciser ce qu'on ignore ».

§ 2. DÉFINITIONS CONFONDANT L'OUTIL ET LE DOMAINE OU ON L'APPLIQUE

Elles font déborder la statistique sur un domaine qui n'est pas le sien. On les retrouve d'abord chez les premiers « apôtres » de la statistique — et l'historique ci-dessus fait comprendre leur confusion avec la recherche économique ou même politique. Achenwall (1749) écrivait : « La statistique est la connaissance approfondie de la situation respective et comparative des Etats. » Sinclair (1785) : « La statistique a pour but de constater la somme de bonheur dont jouit une population et les moyens de l'augmenter. » Jusqu'à la fin du XIX^e siècle les définitions de ce genre restent en honneur.

Cette confusion, nous la retrouvons chez certains statisticiens modernes (et plus spécialement outre-Atlantique) qui ont tendance à absorber toute l'économie politique dans la statistique. Selon eux, tout ce qu'on sait en économie, on le sait sous forme statistique et il est inutile d'envisager une étude économique autonome. C'est une position excessive, estime M. Aftalion, car tout n'est pas mesurable et les autres méthodes d'observation,

en particulier celles qui porteront sur des données non numériques, sont extrêmement utiles.

Avec bon sens, Levasseur reconnaissait à la fin du XIX^e siècle que la « statistique est la lumière de l'économie politique », mais qu'elle n'est pas l'économie politique proprement dite.

Mais on a vu qu'elle était la lumière de bien d'autre chose encore. Et il reste à exprimer dans une bonne définition les caractéristiques propres à cette « lumière ».

§ 3. DÉFINITIONS LIMITÉES

On a proposé la « science des nombres », la science des mesures, la science des moyennes. C'est rester en chemin et s'en tenir aux tâches les plus élémentaires du statisticien. Il est bien évident qu'aussi longtemps que la statistique sera considérée comme une accumulation de données, plus ou moins ordonnées et dégrossies, elle ne méritera pas le nom de science. Elle n'est qu'une méthode — de nutrition — pourrions-nous dire, d'un économiste curieux, qui, lui, étant à la recherche des lois, construira « sa » science en remerciant son fournisseur. Cette position était valable en 1880. Geddes voit dans la statistique *un relevé quantitatif des faits observés dans toutes les branches des sciences* : premier progrès puisque la notation est faite de l'ubiquité statistique et de son emploi généralisé dans toutes les branches de la connaissance. Mais l'effort est insuffisant pour préciser l'originalité de son apport.

Un second progrès se trouve dans les définitions d'A. Aftalion et de Dugé de Bernonville : A. Aftalion (1929 : « Etude numérique des faits et de leurs rapports ») ; Dugé de Bernonville (1935 : « L'ensemble des méthodes ayant pour but l'étude numérique des collectivités, des groupes de faits de même nature, quels que soient d'ailleurs les faits : êtres, choses, ou phénomènes les plus divers ») — puisque l'un introduit l'idée de *rappports* entre les faits et le second l'idée de collectivité ou d'*ensembles*.

La définition donnée par Willcox lui-même (pour remplacer les 115 définitions qu'il avait collectionnées) est de la même veine : la statistique est « l'étude numérique des groupes ou des masses, par l'étude des unités qui les composent, que ces unités soient des hommes ou non, des êtres animés ou inanimés ».

§ 4. DÉFINITIONS MODERNES

La définition de Julin (1921) semble plus évocatrice : « Une méthode qui, par le relevé en masse et l'expression numérique de ses résultats, arrive à la description des phénomènes collectifs et permet de reconnaître ce qu'ils présentent de permanent et de régulier dans leur variété, comme de variable

dans leur apparente uniformité. » Elle soulève cependant les objections des statisticiens puristes, dans la mesure où elle recourt aux notions de *phénomène collectif* et de *relevé en masse* qui sont contestables.

Kendall échappe à ces critiques lorsqu'il écrit : « La statistique est la branche de la méthode scientifique qui traite des données obtenues en comptant ou en mesurant les propriétés de populations de phénomènes naturels. »

Si cette définition a le mérite, par le mot « naturel », d'englober les phénomènes humains, physiques, astronomiques, etc., elle a encore l'inconvénient de ne pas dire comment elle *traite* ces données. C'est pourquoi elle doit être complétée par celle de A. Vessereau qui englobe dans les méthodes statistiques *toutes les recherches dans lesquelles le grand nombre et l'enchevêtrement des facteurs exigent une technique d'interprétation basée sur la connaissance des lois du hasard.*

Résumons donc quelques caractéristiques essentielles de la statistique :

1) Méthode *générale* qui relie des domaines très divers des sciences. Selon l'expression de L. March, elle est bien une « langue commune » ;

2) Elle porte sur des *ensembles* et sur leurs relations ;

3) Elle donne des conclusions probables sur les ensembles *imparfaitement connus* ;

4) Elle aboutit à des lois qui sont des *propriétés de groupe* ;

5) Elle se distingue des mathématiques, science du certain, alors qu'elle est *science de l'incertain* (Kendall), et elle cherche à établir les *limites* de l'incertitude ;

6) Elle s'appuie sur les *lois du hasard*, et elle exploite à fond les quelques domaines où le hasard « organisé » est roi : étude des erreurs — étude des échantillons.

Les résultats acquis, autant que les techniques employées, entraînent une révision des idées sur la Science, comme l'écrivait Henri Berr, qui envisageait une sorte de déterminisme relativisé : « La Science, sans rien perdre de sa grandeur et de sa majesté, consent à cesser de vivre chez les dieux et prend désormais un caractère plus humain. »

BIBLIOGRAPHIE

Historique

Voir ouvrages généraux indiqués ci-dessus, et :

- M. HUBER, La statistique (son histoire, son organisation), dans la *Statistique* (7^e Semaine de Synthèse), 1936. Dans le même ouvrage : PICANIOL, La statistique en histoire ancienne ; ESMONIN, La statistique en histoire moderne.
- R. TATON, *Histoire du calcul*, coll. « Que sais-je ? », n^o 198.
Histoire générale des Sciences (3 tomes parus), Paris, Presses Universitaires de France.

Philosophie de la démarche statistique

- G. DARMOIS, La statistique, *Rev. Enseign. supérieur*, janvier 1960.
- K. ARROW, *La statistique et la politique économique*, Congrès d'Econométrie, 1956.
- C. GINI, Logic in Statistics, *Metron.*, 31-VII-1958.
— Mathematics in Statistics, *Metron.*, 31-VII-1959.
- M. G. KENDALL, The statistical approach, *Economica*, 1950.
- W. F. WILLCOX, Definitions of statistics, *Rev. Inst. int. Stat.*, 1935.
- TORRENS-IBERN, Qu'est-ce que la statistique ?, *J. Soc. Stat.*, Paris, 1957.
- A. SAUVY, L'art de préciser ce qu'on ignore, *J. Soc. Stat.*, Paris, 1952.
- E. BOREL, *Probabilité et certitude*, « Que sais-je ? », n^o 445.

Calcul des probabilités

- B. V. GNEDENKO et A. I. KHINTCHINE, *Introduction à la théorie des probabilités*, trad. du russe, Paris, 1960.
- M. GIRAULT, *Calcul des probabilités en vue des applications*, Paris, 1960.
- A. RENYI, *Calcul des probabilités*, Paris, 1965.
- A. M. YAGLOM et I. M. YAGLOM, *Probabilité et information*, Paris, 1968.

Sur statistique et informatique

- La revue *Informatique et gestion*, Paris.
- J. SHISKIN, *Calculateurs électroniques et indicateurs de conjoncture*, Bordeaux, 1960.
- J. SHISKIN, *Les signaux avertisseurs des récessions et des reprises*, Paris, 1964.
- J. BUREAU, *L'ère logique*, Paris, 1969.

CHAPITRE II

LES APPLICATIONS DE LA STATISTIQUE ET DU CALCUL DES PROBABILITÉS

Lorsque le chevalier de Méré posa à Pascal quelques problèmes relatifs aux jeux de hasard, il n'imaginait certainement pas à quels prolongements et à quelles applications se prêteraient les principes essentiels que son ami allait ainsi avoir l'occasion de découvrir. Il serait difficile aujourd'hui de citer une branche des sciences (physique, chimie, biologie, sciences sociales), une technique quelconque ou même un aspect banal de la vie quotidienne qui ne soit, d'une façon ou d'une autre, concerné par les développements qu'a connus, surtout au xx^e siècle, la science du hasard.

Sans doute même n'est-il pas exagéré de dire que c'est l'ensemble des connaissances humaines, la philosophie de ses limites, de sa construction et de sa démarche, qui prennent dans cette optique nouvelle un jour bien différent de ce qu'il était il y a quelques décades.

Dans ce chapitre, il s'agit seulement de fixer les idées sur les propositions qui viennent d'être énoncées. La section I dégage quelques notions fondamentales : mesure, échantillon, être scientifique, et énumère les tâches possibles de la statistique. La section II est consacrée à la présentation sommaire des applications dans les différentes branches de la science et de l'activité — une étude exhaustive, si elle est possible, sortirait du cadre de cet ouvrage ; la section III, aux problèmes posés par l'emploi de la statistique dans les sciences sociales. La section IV s'efforce enfin de dégager l'essentiel de la démarche du statisticien moderne.

SECTION I. — Le champ de la statistique

Au début de toute démarche scientifique ou technique il y a l'*identification* de telle qualité ou de telle quantité ; une action de repérage sinon de mesure s'impose : par sa première intervention, la statistique permet de savoir de quoi on traite. Mais comme les observations sont en général limitées en nombre, il faut savoir comment on peut *généraliser*, passer de la partie au tout, de l'échantillon à la population totale : la seconde tâche est celle de l'*estimation*.

Quand on confronte plusieurs populations ou des populations et des échantillons qui en sont tirés (les populations sont ici des ensembles d'hommes ou d'objets, de qualités ou de quantités relatives à certains caractères), il faut pouvoir éprouver la *validité des comparaisons*. Rechercher si les différences ou les similitudes sont significatives, c'est faire des tests de signification, troisième tâche de la statistique.

Juger d'une évolution, d'un phénomène, *dégager une tendance* cachée par d'innombrables fluctuations à court terme et par des variations accidentelles, c'est encore un traitement statistique important. Le travail sur séries temporelles, le dégagement des évolutions chronologiques est une quatrième tâche.

Enfin la *recherche des liaisons* entre deux ou plusieurs phénomènes (cinquième tâche) — par exemple longueur et température pour le physicien, quantité de pluie et rendement du blé pour l'agronome, la température et les rhumes de cerveau pour le médecin, vitesse des voitures et accidents pour l'assureur, industrialisation et niveau de revenu pour l'économiste — implique des travaux statistiques délicats dont la technique, on l'imagine déjà, est toujours à peu près la même quel que soit le domaine d'application.

Il n'est pas sans intérêt de revenir sur quelques-uns de ces types de travaux.

§ 1. IDENTIFICATION. PROBLÈMES DE MESURE

a) LA MESURE RÉPÉTÉE DANS LE TEMPS. — Prenons l'exemple de la mesure d'une longueur.

Si je mesure la pièce où je me trouve, deux fois de suite, avec un double décimètre, il y a peu de chances que je trouve deux fois le même résultat. La longueur de la pièce ne peut pas être représentée par UN nombre (526 ou 527 cm) ; elle ne peut guère être définie, si je recommence un grand nombre de fois l'opération de mesure, que par une figure en escalier, un *histogramme* (voir chap. II, titre IV, graphique p. 270) portant en abscisse les mesures exprimées en centimètres et en ordonnée la fréquence relative de ces mesures.

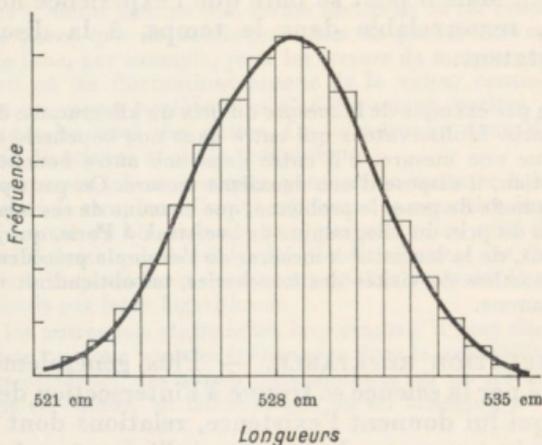


FIG. 1

Dire : la pièce mesure 526 cm est une imposture.

Dire : d'après les mesures que je viens de faire, la pièce mesure en moyenne 527 cm, c'est de l'arpentage scrupuleux.

Si on ne veut retenir qu'un seul nombre, une seule « statistique », ce sera en sous-entendant une phrase comme celle-ci : il y a tant de chances pour que l'erreur que je commets en donnant ce nombre soit inférieure à telle valeur. La science du hasard peut permettre de chiffrer une telle proposition, eu égard aux conditions de l'expérience, au nombre et à la qualité des mesures effectuées. Ce sera, par exemple : longueur de la pièce : 528 cm avec 95 chances sur 100 pour que l'erreur que je commets soit inférieure à 3 cm.

Ainsi toute connaissance expérimentale n'a, au départ, de sens qu'à l'intérieur d'une théorie probabiliste. On conçoit alors que toute science expérimentale dépende essentiellement de la science du hasard.

b) LA MESURE RÉPÉTÉE DANS L'ESPACE. — L'exemple précédent est celui d'une expérience répétable *in vitro*, c'est-à-dire d'une expérience dont le responsable est à même de contrôler les conditions et qu'il peut renouveler à volonté. Ce type d'expérience est celui de l'expérience de mécanique (par exemple, mesure de la longueur d'un ressort en fonction de la charge). Mais il peut se faire que l'expérience ne soit pas, par nature, renouvelable dans le temps, à la discrétion de l'expérimentateur.

C'est le cas par exemple de la mesure du prix du kilogramme de beefsteak au détail à Paris. L'observateur qui entre dans une boucherie et demande ce prix effectue une mesure. S'il entre dans une autre boucherie et pose la même question, il dispose d'une deuxième mesure. On peut penser, c'est une façon commode de poser le problème, que chacune de ces mesures donne une estimation du prix du kilogramme de beefsteak à Paris, qui joue le rôle, en plus abstrait, de la longueur à mesurer de l'exemple précédent. Au bout d'un certain nombre de visites des boucheries, on obtiendrait d'ailleurs le même histogramme.

c) LA RELATION RÉPÉTABLE. — Plus généralement, tout être examiné par la science se trouve à l'intersection de relations répétables qui lui donnent l'existence, relations dont la finesse dépend des instruments de mesure utilisés et qui, pour un instrument suffisamment fin, ne sont plus que des relations *stochastiques*, c'est-à-dire qu'elles relient des histogrammes (comme celui de la fig. 1) et non des nombres.

d) PERMANENCES STATISTIQUES ET LOIS STATISTIQUES. — Jusqu'ici nous avons envisagé quelques « populations de mesures » (longueurs ou prix). Mais les séries statistiques, dans tous les domaines, nous livrent d'autres populations dont les *régularités* sont à découvrir. Le taux de masculinité (rapport des naissances de garçons aux naissances totales) est voisin de 0,51 ; les observations faites dans différents pays et pour des races différentes permettent d'établir un histogramme analogue à celui de la fig. 1, avec une très faible dispersion autour de la valeur centrale 0,51. La taille des conscrits, les points d'impact des balles tirées par un fusil tenu par un étau fixe et bien d'autres phénomènes vont nous montrer des distributions semblables. Très tôt, on a été amené à conclure que de telles

distributions de populations (conscrits, points d'impact, etc.) obéissaient à certaines régularités dont l'expression est possible : tel est le but des lois statistiques.

Il y a de très nombreuses lois statistiques. Le triangle de Pascal trouve son aboutissement dans la *loi binomiale*. La *courbe en cloche*, connue de tous, est la représentation de la *loi normale* (ou loi de Laplace-Gauss) qui est apparue d'abord comme une forme limite de la loi binomiale. Elle est importante parce que de nombreuses lois de probabilité convergent vers elle. Et elle joue, par exemple, pour les *erreurs* de mesure et de façon générale partout où les fluctuations autour de la valeur centrale sont dues à des erreurs ou causes accidentelles s'additionnant, petites, nombreuses et indépendantes les unes des autres. On la retrouve donc dans toutes les branches de la science et de la technique.

De nombreuses lois statistiques peuvent se ramener à la *loi normale* : il suffit parfois de remplacer l'unité exprimant le caractère étudié (longueurs, tailles, rendement du blé, revenus, etc.) par une autre unité (inverse, racine, carré, cube, logarithme). Ainsi les lois de distribution des revenus qui sont fréquemment dissymétriques deviennent des lois normales lorsque les revenus sont remplacés par leurs logarithmes.

Parmi les autres lois statistiques importantes, il faut citer également la *loi de Poisson* qui joue pour les événements dont la probabilité est très faible (études des accidents, des appels téléphoniques sur un réseau, des solutions très diluées en chimie, en bactériologie ou en biologie).

e) LA RÉDUCTION DES DONNÉES. — Les lois les plus couramment utilisées ont une forme analytique connue ; on peut alors valablement remplacer l'ensemble des données par quelques grandeurs caractéristiques (voir titre V, chap. I^{er}). Les meilleures sont celles qui définissent la position et la dispersion des données et qui sont les paramètres dont la loi dépend.

Les *paramètres de position* caractérisent la valeur centrale des données (moyenne, médiane par exemple) et les paramètres de dispersion précisent les écarts entre elles ou entre elles et une valeur typique comme la moyenne. On verra par exemple plus tard que, pour une distribution normale, la *moyenne arithmétique* et une unité de dispersion appelée *l'écart-type* sont les paramètres qui permettent d'exprimer la loi — et donnent à eux seuls une information très riche et très suffisante sur l'ensemble des données.

§ 2. ESTIMATION. DE L'ÉCHANTILLON A LA POPULATION

Tout observateur n'effectue jamais qu'un nombre limité d'observations et cet ensemble d'observations qu'il effectue n'est jamais qu'un *échantillon* parmi la population infiniment grande de séries d'observations qu'il conviendrait de faire pour atteindre la réalité.

Une opinion répandue chez les non-spécialistes est que les théories de l'échantillonnage ne s'appliquent que dans le cas du sondage, volontairement limité à une partie de la population (d'hommes ou d'objets). Il faut généraliser et dire que l'observateur n'atteint jamais que des sous-populations desquelles il a l'ambition d'induire des relations valables au niveau de la population entière.

Même si on interroge les 46 millions de Français, on n'atteint qu'une sous-population de la vraie population française — ou d'une « population parente » — qui, seule, est justiciable des lois statistiques. Ce passage de l'échantillon à la population est le passage de la mesure expérimentale à la mesure vraie, de l'être contingent à l'être scientifique, de la relation expérimentale à la loi. C'est une démarche essentielle de l'activité scientifique qui consiste à calculer une *estimation* ou valeur approchée pour une population globale dont on ne connaît qu'une partie : on détermine des *intervalles de confiance* dans lesquels les vraies valeurs à estimer ont une forte probabilité de se trouver.

§ 3. LES TESTS D'HYPOTHÈSE

Ils représentent une branche décisive de l'analyse statistique ; leurs techniques progressent très rapidement et leur champ d'application est immense. « Une certaine hypothèse concernant une ou plusieurs populations ou une ou plusieurs lois de distribution ayant été formulée, l'utilisation d'un test statistique approprié permet de déterminer dans quelle mesure cette hypothèse apparaît *infirmée* par l'information apportée par un ou plusieurs échantillons, c'est-à-dire dans quelle mesure cette information peut être considérée comme incompatible avec l'exactitude de l'hypothèse formulée. Si cette mesure

est forte, l'hypothèse est rejetée sur la foi du test ; si elle est faible, on conclut seulement à l'absence de raisons d'évidence suffisante pour la rejeter : l'hypothèse est alors acceptée, mais seulement de façon provisoire. Les tests statistiques permettent ainsi de *fonder l'induction* en donnant le moyen, à partir de l'information nécessairement incomplète fournie par des échantillons, de formuler certains jugements concernant les populations où ils sont prélevés » (A. Laurent). On trouvera des exemples de tests dans la section suivante de ce chapitre, au titre V, et dans notre ouvrage *Statistique et observation économique*.

§ 4. ÉTUDE DES ÉVOLUTIONS CHRONOLOGIQUES

Les moyens dont la statistique dispose pour étudier des données échelonnées dans le temps sont également nombreux et efficaces. Dégager une tendance à long terme, effacer les perturbations qui cachent le mouvement général, séparer divers mouvements composants (dont, en économique, les variations saisonnières, la fluctuation cyclique et la croissance longue sont des exemples) sont des tâches qui peuvent être menées avec des outils plus ou moins précis et plus ou moins compliqués. Contrairement à une opinion fréquente, ces études ne sont pas exclusivement limitées au domaine économique : on les trouve dans toutes les sciences de l'homme et de la nature, par exemple en médecine — où les maladies ont à la fois leurs pointes saisonnières et de véritables cycles qui s'étendent sur plusieurs années — ou en météorologie.

§ 5. RECHERCHE DES LIAISONS ENTRE PHÉNOMÈNES

Partant de la liaison fonctionnelle entre deux variables (y étant expliqué par x et à toute valeur de x correspondant une valeur précise de y), on s'aperçoit, dans tous les domaines de recherche, soit que l'observation de ces variables n'a pu être conduite de façon parfaite, c'est-à-dire qu'elle est affectée d'erreurs de mesure, soit que la liaison établie n'est valable que dans les grandes lignes, c'est-à-dire qu'elle est affectée par des facteurs incontrôlables, par de nombreuses causes de varia-

bilité. Le but des méthodes statistiques (corrélation, régression) est de mesurer l'ampleur de ces perturbations, et éventuellement de « filtrer » les phénomènes étudiés pour que la nature et la solidité de leur liaison apparaissent au grand jour.

Il est certain par exemple que la consommation d'essence dans des unités territoriales données (comme le département) est fonction de l'importance du parc automobile. Mais la liaison est affectée par de multiples autres variables : revenu de la population, répartition par professions, état des routes, structure du commerce de l'essence, mouvements du tourisme, etc. De même la charge de rupture d'un acier dépend de sa teneur en carbone. Mais la liaison peut être plus ou moins troublée par les variations de sa teneur en manganèse et en silicium. Dans tous les domaines, on trouve des cas où l'expérience reste *brute* et où, faute de pouvoir physiquement éliminer les facteurs secondaires ou perturbateurs, on est obligé de ruser et de chercher par le calcul quelle est leur influence *probable*.

La conclusion à tirer, pour cette section méthodologique, c'est qu'on est loin du tableau de Claude Bernard (*Introduction à l'étude de la médecine expérimentale*) qui est resté longtemps un modèle pour les savants. La succession « observation-hypothèse-expérimentation » paraît aujourd'hui moins simple qu'on ne le pensait à la fin du siècle dernier. Non pas certes que l'étude de l'aléatoire entraîne la négation des moyens élaborés par le positivisme. C'est plutôt, dans la même ligne, un élargissement de la recherche, la découverte de nouveaux outils au nom d'une inquiétude. Seul un certain déterminisme tranquille est condamné — et dans les grandes lignes objectifs de la connaissance et méthodes restent les mêmes. La méthode dite expérimentale (qui consiste à isoler certains facteurs et à laisser varier quelques-uns d'entre eux en laissant les autres constants) reconnaît maintenant ses limites : son application rigoureuse, dit A. Laurent, « exige la réalisation de conditions susceptibles de caractériser seulement des expériences purement idéales qu'on peut regarder tout au plus, quand elles constituent autre chose que de simples abstractions dépourvues de caractère opérationnel, comme des *cas limites* d'expériences réalisables ». La méthode statistique reprend un contact plus réaliste, à la fois *plus timide* — puisqu'elle reconnaît que l'observateur ne peut prendre les observations que comme elles sont et comme elles viennent, dans un écheveau embrouillé de para-

mètres et de variations de facteurs non contrôlées — et *plus ambitieuse* aussi puisqu'elle s'attache à réduire l'incertitude des conclusions tenant à l'imprécision des mesures, à l'impureté des observations, à l'information fragmentaire, à l'enchevêtrement des relations. Sans la science du hasard, l'*induction*, ce « procédé » logique hasardeux qui consiste, dit encore A. Laurent, à « tirer des conclusions d'une portée générale à partir d'informations incomplètes et particulières », n'aurait qu'une efficacité limitée. La statistique lui rend toute sa puissance, en mesurant l'incertitude des conclusions inductives, en contrôlant cette incertitude et même, pourrait-on dire dans une image assez grossière, en domestiquant l'aléatoire.

SECTION II. — Les applications statistiques dans les sciences et les techniques

Si, il y a quelques décades, on pouvait chercher des exemples d'application de la statistique probabiliste dans différents domaines, l'expansion de la méthode a été si large et si rapide qu'aujourd'hui on serait plus bref en cherchant les branches où elle ne sert pas.

Tenant pour acquis cet aspect essentiel de la présence du hasard dans l'activité scientifique et dans l'activité humaine en général, nous donnerons, pour fixer les idées et non pour faire une peinture complète, quelques exemples pris dans des domaines très éloignés les uns des autres.

§ 1. LA SCIENCE PHYSIQUE

La science physique ne prétend plus que deux êtres physiques de même définition (deux gouttes d'eau) sont identiques. Cette identité ne dépend jamais que de l'échelle de l'observation. L'observation à une échelle plus fine permet de découvrir de nouveaux phénomènes dont la valeur explicative est plus grande.

a) THÉORIE CINÉTIQUE DES GAZ. LOI DE MAXWELL ET MÉCANIQUE STATISTIQUE. — La théorie cinétique des gaz conduit à considérer qu'un gaz à la pression ordinaire se compose d'un très grand nombre de molécules. On aura une idée de ce nombre si on sait que, pour compter, à raison de 1 000

par seconde, les molécules contenues dans un millimètre cube d'un gaz à 0 °C sous 760 mm de mercure, il faudrait 9 000 siècles.

Imaginons maintenant un récipient de 22,4 l contenant $22,4 \times 10^6$ fois le nombre précédent de molécules, et divisons-le par la pensée en 2 récipients, A et B, égaux, de 11,2 l chacun. Le calcul des probabilités enseigne que, si on tire au hasard la place de chacune des molécules (dans A ou dans B), il y a une chance sur 10 pour que le nombre des molécules situées en A dépasse de 100 milliards celui des molécules situées en B. Ces 100 milliards ne représentent que la 100 milliardième partie du nombre total des molécules et, si cette éventualité se produisait réellement, elle serait impossible à repérer par les méthodes actuellement connues.

Les fluctuations de densité ne deviennent observables que dans l'étude des gaz très raréfiés. L'observation confirme alors le calcul : on a constaté effectivement, dans ce cas, des fluctuations — par exemple dans l'étude de certains phénomènes d'opalescence.

Ainsi, à partir de l'hypothèse de l'inorganisation des molécules, devient-il possible d'expliquer des phénomènes constatés à une autre échelle (mesure des pressions et des densités). La loi de Maxwell donne la répartition des vitesses des diverses molécules d'une masse gazeuse.

La théorie cinétique des gaz n'est qu'un cas particulier de la mécanique statistique (Gibbs) dont les applications à la physique sont devenues de plus en plus nombreuses depuis que le discontinu remplace le continu (discontinuité de la molécule, de l'énergie, de l'électricité). Un des derniers succès de ce point de vue est dû à l'élaboration du corps de théories connu sous le nom de mécanique ondulatoire grâce auquel ont pu être dépassés des conflits logiques apparemment insolubles comme celui concernant la structure de la lumière.

b) ASTRONOMIE. — La distribution des étoiles dans l'espace et les lois de répartition de leurs vitesses ont été étudiées par des méthodes analogues à celles de la théorie cinétique des gaz. On est là en face d'une expérience unique, d'un seul échantillon, et l'induction jusqu'à l'hypothèse cosmogonique (origine, structure et avenir de l'Univers) ne saurait prétendre à la rigueur de disciplines qui permettent d'effectuer des opérations répétables.

Néanmoins on peut dire que les corrélations de la statistique stellaire, les analyses statistiques du spectre des étoiles, l'étude de la nature des étoiles doubles, de la fluctuation de luminosité des Céphéides, des amas d'étoiles, de l'éloignement des nébuleuses, etc., ont engendré pour l'astronomie des progrès plus rapides que ceux dus aux télescopes des observatoires les mieux outillés.

Autre exemple d'application statistique : l'étude de l'origine des comètes. Le calcul des probabilités permet d'affirmer que, pour les comètes, les orbites elliptiques ont beaucoup moins de chances de se produire, sauf

hypothèses exceptionnelles, que les hyperboliques. L'observation d'un nombre anormalement élevé d'orbites elliptiques conduit donc à présumer ces hypothèses exceptionnelles — et à admettre en fait qu'il s'agit de corps faisant partie intégrante du système solaire et n'ayant donc pas d'origine cosmique.

§ 2. SCIENCE BIOLOGIQUE

a) GÉNÉTIQUE. — C'est G. Mendel qui, vers 1860, a établi les lois d'hérédité des caractères à la suite d'observations statistiques sur les croisements de diverses variétés de pois. La génétique s'est ensuite rapidement développée grâce à des travaux statistiques sur des animaux à développement et à reproduction rapide (par exemple la mouche du vinaigre et la souris).

« A un moment où rien n'était connu du mécanisme de l'hérédité, ni de la structure des cellules germinales, il a été possible de démontrer par cette méthode l'existence de l'hérédité, d'en mesurer l'intensité, et cela chez l'homme, c'est-à-dire sur un organisme où l'étude expérimentale des croisements ne peut se faire » (R. A. Fisher). En croisant des individus de gènes différents, on a pu étudier comment les gènes se séparent et se regroupent dans la succession des générations. Il a fallu utiliser des calculs de corrélation et des tests statistiques. C'est ainsi que K. Pearson a montré l'hérédité des caractéristiques physiques (tailles) et mentales par étude des mesures du père, de la mère, du fils et de la fille, puis par étude des mesures frère-frère, sœur-sœur, frère-sœur et sœur-frère...

Il est intéressant d'ajouter que lorsque l'étude des cellules et des chromosomes a été assez avancée on a pu confirmer les hypothèses faites auparavant au moyen des calculs statistiques.

Autre exemple : l'homogamie (similitude de certaines caractéristiques chez le mâle et la femelle dans le couple) a pu être étudiée statistiquement : les coefficients d'homogamie sont plus forts pour les invertébrés que pour les autres animaux. Ils ont une importance non négligeable pour les humains. On a également pu voir l'influence de l'homogamie sur la tendance à conserver les caractères et à créer des sous-groupes ainsi que la relation étroite entre elle et la fertilité.

b) SÉLECTION. — Elle est devenue réalisable grâce à la statistique. Si on considère un élément mesurable chez tous les individus d'une même espèce, on peut dresser la statistique des valeurs prises par cet élément dans un échantillon de la population totale.

Pour la taille, par exemple, on constate que les Français de 20 ans se répartissent selon une courbe de distribution normale (ou gaussienne). On peut donc dire que tel individu est caractérisé par un certain écart par rapport à l'individu « moyen » : c'était, on l'a vu, le point de vue de Quételet.

Si on fait la proposition inverse, que la constatation d'une distribution gaussienne dans un échantillon pris au hasard et suffisamment représentatif d'une population donnée, est la marque de l'homogénéité de cette population, on fournit au biologiste un procédé lui permettant de reconnaître si un ensemble d'individus est ou non homogène — et par suite le moyen de sélectionner des races pures, ce qui est de première importance, notamment en agronomie.

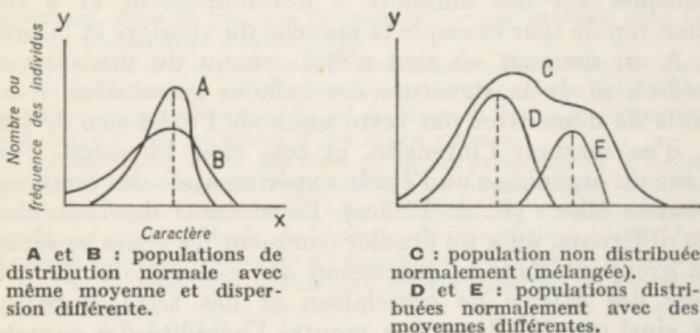


FIG. 2. — Types de « populations »

Le recours aux lois de croisements rend la connaissance opératoire : on peut associer dans une génération les caractéristiques (taille, fertilité, résistance) des parents — et les maintenir stabilisées dans une lignée pure : les caractéristiques sont alors, pour les « individus » des générations successives, distribuées normalement — mais ces différences entre individus restent constantes ; c'est seulement si on constate que la moyenne et la dispersion de cette distribution normale changent ou que la distribution cesse d'être gaussienne que l'hypothèse d'un mélange de variétés doit être faite.

c) DE LA BIOMÉTRIE HUMAINE A LA POLITIQUE DE LA SANTÉ PUBLIQUE ET A LA RATIONALISATION DES PRODUITS DE CONSOMMATION. — Les mesures et pesées effectuées sur des échantillons de population donnent depuis quelques années des

indications précieuses. A la fin de la guerre, les restrictions alimentaires ont eu des conséquences sur la santé publique et plus particulièrement sur la situation physiologique des enfants.

En France, l'Institut National de la Statistique et des Etudes Economiques et la Direction de l'Hygiène Scolaire ont effectué des sondages sur les dossiers médicaux de 14 000 écoliers (1938, 1942, 1946). Des enquêtes plus vastes en 1950 (300 000 observations) et en 1955 permirent d'établir des tables d'étalonnage (tailles et poids) ainsi que la combinaison de ces deux données, et de faire ressortir d'importantes différences entre départements. On a tiré des travaux de ce genre d'intéressantes conclusions sur la protection alimentaire des écoliers (Dr J. Sutter), sur la croissance et la stature des enfants (J. Tremolières), sur les âges d'éruption dentaire, etc.

De nombreux chercheurs se préoccupent également de l'augmentation de la taille humaine (1 cm en 10 ans en Europe occidentale et en Amérique du Nord) et de l'augmentation du poids moyen des nouveaux-nés (5 % entre 1900 et 1940 en Allemagne et en Suisse — d'après étude de 200 000 naissances en clinique). On a essayé de dégager l'influence des conditions de vie (amélioration de l'hygiène, de l'équilibre alimentaire, stimulation des techniques nouvelles, prolongation artificielle du jour, etc.). Mais en 1961, le Dr R. Nold a publié les résultats d'une longue étude statistique montrant que le phénomène tient essentiellement à la plus grande mobilité humaine. On savait déjà que, avec des croisements d'animaux ou de végétaux qui se distinguent par un assez grand nombre de dispositions héréditaires, la descendance dépasse fréquemment la taille des ascendants. La croissance est accélérée et les besoins de nourriture accrus. Cette « luxuriance des sujets hybrides » se vérifie chez les humains. Autrefois les mariages se contractaient entre personnes vivant dans un petit rayon, tandis qu'aujourd'hui les lieux d'origine des époux sont souvent séparés par de grandes distances. Dans l'enquête du Dr Nold, il apparaît que les « grands » enfants sont issus de mariages entre personnes dont les origines sont séparées par de grandes distances et les « petits » de mariages entre voisins. Aux Etats-Unis, la fusion des races ajoute encore quelques changements : diminution du nombre des individus blonds et régression de certaines maladies héréditaires. L'industrialisation et les déplacements de masses ont donc des conséquences biologiques non repérables sans le secours des techniques statistiques.

Mais les progrès de l'étude statistique du corps humain ont permis d'autres applications, notamment dans la fabrication des articles d'habillement, des gants, des chaussures, etc. Sans de telles études, les produits de confection ne correspondent pas à la « demande », et les *soldes* offerts périodiquement par les magasins, dans lesquels on ne trouve guère que des articles trop grands ou trop petits, en témoignent. Pour

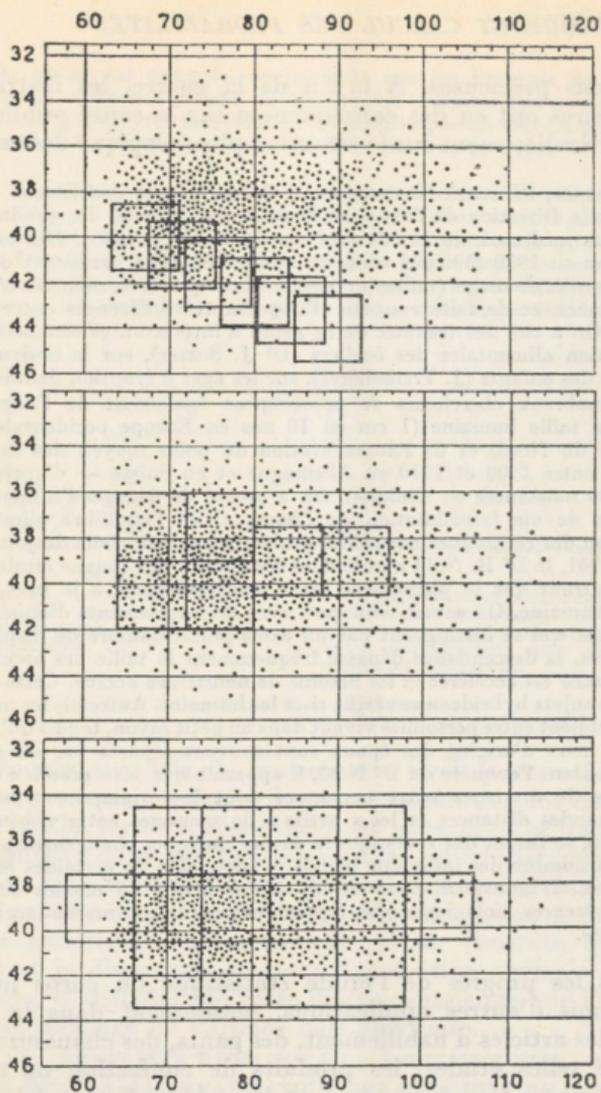


FIG. 3. — Détermination de la « juste mesure » dans la confection féminine
 (Source : G. TAGLIACARNE, *Tecnica et Pratica delle Ricerche di Mercato.*)

ajuster la production à la répartition exacte des mensurations de la clientèle, sans augmenter de façon considérable le nombre des modèles, il a fallu procéder à des mesures des tailles (taille, tour de hanches, tour de poitrine, longueur du dos, des jambes, des bras, etc.), rechercher celles qui sont au plus haut degré représentatives de toutes les autres. Des calculs de corrélation et une analyse statistique poussée ont donné des résultats remarquables, notamment aux États-Unis, en Hollande, en France (E. Morice) et en Italie (G. Tagliacarne).

La figure 3 donne l'exemple des résultats obtenus en Hollande par l'étude du grand magasin *De Bijenkorf* : 14 caractères anthropométriques ont été mesurés sur 5 000 femmes et ont donné les corrélations les plus fréquentes, utilisables pour la fabrication en série. Une de ces corrélations concerne le tour de taille et la longueur du dos. G. Tagliacarne fait le commentaire suivant : « En reportant sur un graphique les mesures du tour de taille et de la longueur du dos, on se rend compte de la distribution caractéristique de ces mesures. Les points compris dans chaque carré du graphique indiquent combien de Hollandaises sur 1 000 ont ces mesures et donc peuvent trouver des vêtements qui leur vont.

« Le graphique supérieur illustre la solution adoptée avant l'étude des

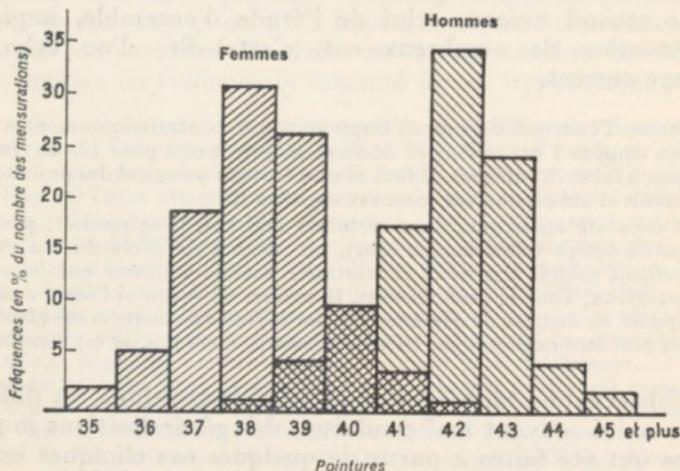


FIG. 4. — Sondage sur la répartition optimum des pointures de chaussures (France, 1953)
(C.E.M.P., *Le marché français de la chaussure.*)

Etablissements De Bijenkorf : avec 6 tailles (les 6 carrés) on ne pouvait satisfaire que 27 % des clientes. Après l'étude (graphique central), 6 tailles mieux choisies permettent de satisfaire 66 % des femmes et 14 tailles (graphique inférieur) correspondent à 91 % des demandes probables. » Ainsi on peut fabriquer sans grands risques d'erreur le nombre voulu de vêtements de chaque taille (élimination des pertes d'invendus) et on évite les retouches (élimination des coûts de main-d'œuvre hors usine).

Autre exemple : une étude du Centre d'Etude de la Productivité menée sur 19 250 personnes en 1953 a indiqué la distribution des pointures. Pour les hommes, la taille la plus fréquente est le 42, pour les femmes le 38. La dispersion est plus grande pour les femmes que pour les hommes.

d) MÉDECINE. — Bien que les médecins aient été longtemps réticents à l'égard de la statistique scientifique, aucun domaine n'a plus besoin d'une observation rigoureuse, testée et interprétée des phénomènes, et ceci à un double niveau :

Pour le *diagnostic* (1^{er} niveau), le praticien doit tenir compte de la fréquence relative des symptômes qui apparaissent, seuls ou associés, dans certaines maladies : une situation individuelle se définit par une combinaison de caractères, à établir et à confronter ensuite avec les « distributions globales ».

Le second niveau, celui de l'étude d'ensemble, implique l'observation de nombreux cas, c'est-à-dire d'un échantillonnage correct.

Prenons l'exemple du *cancer*. Supposons que la statistique montre qu'à Paris on compte 1 cancer sur 15 décès et en province 1 pour 18. Ce résultat n'est pas, à lui seul, suffisant. Il faut encore recourir au calcul des probabilités pour savoir si oui ou non cet écart est significatif.

On constate aussi que, dans certaines maisons, il se produit, pendant un laps de temps relativement court, un nombre de décès dus au cancer relativement grand. Une telle constatation incline à penser que le cancer est contagieux. Tout le problème est, là encore, de savoir si l'écart constaté par rapport au nombre de décès moyen par maisons de contenance identique peut ou non être expliqué par le jeu du hasard, n'est pas ou est significatif.

L'observation en médecine et en thérapeutique est difficile et elle a été souvent mal conduite : des généralisations imprudentes ont été faites à partir de quelques cas cliniques isolés. Souvent aussi des résultats de travaux portant sur un plus grand nombre de cas ont été assimilés à tort à des résultats de sondages.

C'est ainsi que les expériences faites pour examiner la valeur de la vaccination B.C.G. en prophylaxie antituberculeuse ont fait l'objet de violentes controverses pendant 30 ans. Les résultats (favorables) obtenus par observation des sujets vaccinés sur demande ne prouvent rien : en effet, on ne peut pas les comparer avec les sujets qui ne se font pas vacciner : il y a entre les 2 groupes des différences décisives de groupe social, de structure familiale, de genre de vie, etc. Les discussions n'ont cessé qu'avec des travaux irréprochables comme celui du *Medical Research Council* de Grande-Bretagne : échantillon au hasard de 56 000 enfants dans 3 grandes villes ; subdivision en groupes à réaction pré-tuberculinique positive et négative, puis en vaccinés (2 vaccins différents) et en non-vaccinés ; calcul des intervalles de confiance, etc. Les cas de tuberculose sont apparus 5 fois plus fréquents chez les non-vaccinés.

Dans quelques pays, les *médicaments* modernes font l'objet d'une expérimentation thérapeutique contrôlée : division au hasard des sujets en 2 groupes, soumis à 2 traitements différents, et comparaison des 2 résultats par un test de signification. Mais la vraie difficulté réside dans le tirage au hasard et dans la constitution d'un échantillon assez large.

Faute de tirage au hasard, explique D. Schwartz, les médecins ont admis pendant longtemps « que les malades de 2 années successives ou de 2 hôpitaux voisins constituent des groupes comparables. Il a fallu un certain nombre de déboires pour mettre en évidence la fausseté de ces hypothèses.

« Les maladies sont changeantes, tout particulièrement les maladies infectieuses ; c'est cette *variabilité* qui, au moment de l'entrée en scène de l'auréomycine, fit croire à son efficacité dans le traitement de la pneumonie atypique ; il fallut attendre 4 ans, pendant lesquels avaient été utilisés des centaines de kilogrammes d'auréomycine, pour qu'un essai correct en démontrât la parfaite inutilité. Le recrutement même des malades change d'une année à l'autre, d'un hôpital à l'autre. » Ainsi la mortalité par cancer de la langue sembla augmenter rapidement entre 1945 et 1950... parce qu'un centre spécialisé attirait un nombre croissant des cas les plus graves. Un médecin célèbre, un traitement nouveau changeant la structure de la clientèle et la gravité des cas soumis à l'examen et au traitement. La ventilation des malades entre le chirurgien et le médecin peut aussi varier dans le temps ou d'un centre à un autre. Les comparaisons ne sont valables que si on dispose d'un nombre suffisant de malades scientifiquement pris en charge. D. Schwartz donne à titre d'exemple un calcul d'échantillon : avec un risque de déclarer avantageux un traitement qui ne l'est pas et un risque de rejeter un traitement avantageux pris à 5 %, et une proportion de guérisons avec le traitement A estimée à 50 % des malades, l'effectif total à soumettre au traitement A et au traitement B (qu'on veut comparer à A)

est de 1 100. Or, si le plus grand centre anticancéreux français reçoit par an 7 000 nouveaux malades, 400 seulement ont le cancer le plus fréquent (cancer du sein). Sur ce nombre 200 seulement n'ont jamais subi de traitement avant leur entrée. En séparant les cas cliniquement différents et en éliminant les malades pour lesquels les traitements sont contre-indiqués, il reste 30 malades par an sur lesquels on peut mener l'étude. D'où la nécessité de délais très longs ou de méthodes d'investigation plus larges, à l'échelle internationale.

Mais déjà, l'étude des épidémies, des groupes sanguins, la standardisation des produits pharmaceutiques, la connaissance des populations d'insectes et de parasites et de leurs migrations — donc de leur rôle dans la transmission des maladies — ont entraîné des progrès décisifs. La malaria est vaincue au moins autant par l'étude statistique des moustiques que par les médicaments.

Un autre exemple est fourni par l'étude de S. Ledermann (Colloque international d'analyse factorielle, 1955) sur la surmortalité des Français adultes, du sexe masculin. Cherchant les corrélations possibles entre la série statistique des causes de décès par départements et diverses autres séries, il a mis en évidence le rôle de l'alcoolisme. Les corrélations les plus étroites de ce facteur alcoolique concernent la mortalité par maladie du foie (ce qui est évident à cause des cirrhoses alcooliques), tuberculose, hémorragie cérébrale, pneumonie, néphrite, suicide, accident, homicide, toutes corrélations qui n'avaient pas été mesurées mais qui étaient connues ou, au moins, présumées par expérience clinique. Un fait plus nouveau est la mise en évidence d'une liaison entre l'alcoolisation et la mortalité par cancers (digestifs notamment). Les équations de régression obtenues permettent d'estimer la part de mortalité des hommes âgés de 35 à 50 ans imputable à une alcoolisation excessive. Le résultat de l'analyse portant sur les femmes constitue une contre-épreuve valable; de même que la comparaison avec les statistiques de mortalité de pays tels que la Hollande où l'alcoolisme a été considérablement réduit.

Un problème comparable est posé par l'existence d'une forte corrélation linéaire négative entre le *coefficient de natalité d'un pays donné et la consommation par habitant de protéines animales*. Si cette corrélation apporte aux sociologues un certain nombre de réponses, elle pose aux médecins une série de nouveaux problèmes : les travaux sur les pays sous-développés contribuent ainsi à l'enrichissement de la connaissance médicale.

On sait que les Anglais ont donné à cette méthode de décou-

verte, par effet induit, par accident pourrait-on dire, de résultats dans un domaine extérieur à celui dans lequel la recherche avait été entreprise, le nom de *serendipity*. On doit ici remarquer que cette *serendipity* devient, en analyse statistique appliquée, parfaitement logique.

§ 3. LES TECHNIQUES

On ne peut envisager de donner ici tous les éléments qui s'imposeraient.

a) ACTUARIAT, ASSURANCE. — L'exploitation des travaux des démographes — tables de mortalité, lois de survie, etc. — a donné une base scientifique à l'assurance qui est peut-être une des grandes caractéristiques de notre temps.

Pour l'assurance-vie, par exemple, les actuaires calculent, pour une population donnée, quelle est la probabilité pour que — par âges, par professions, par lieux — le nombre des décès ne dépasse pas un certain niveau. Des tables de survie, d'espérance de vie à chaque âge, etc. sont régulièrement établies, mises à jour.

Des techniques analogues sont employées pour les accidents, les accidents du travail et des risques sans cesse plus nombreux (incendie, grêle, inondation, bientôt risque atomique, etc.).

Grâce à ces travaux, le statisticien peut dire par exemple : dans les 365 jours qui viennent, 1 500 Parisiens mourront de mort violente ; le total des victimes est prévisible avec une faible marge d'erreur, mais les noms restent inconnus. On retrouve ici le déterminisme des ensembles et l'incertitude des individus déjà évoqués plus haut.

b) GRANULOMÉTRIE. — L'étude statistique des populations de grains a fait progresser d'innombrables techniques : le *bâtiment* par une meilleure connaissance des propriétés mécaniques du béton qui sont fonction des grains composants ; les *charbonnages*, pour le concassage, le lavage du charbon, pour la sécurité grâce à l'étude des poussières ; les *centrales électriques*, par l'étude des rendements de combustibles formés de différentes tailles de grains ; les *poudreries*, etc.

c) MÉTÉOROLOGIE. — Cette discipline s'est renouvelée grâce aux analyses statistiques modernes. Corrélations entre facteurs, entre lieux, étude précise des évolutions, tests de signification permettent de dégager des relations qui donnent plus de sécurité dans la prévision.

d) AGRONOMIE. — Les liaisons entre facteurs météorologiques et rendement des récoltes donnent un nouvel exemple de recherches de corrélations : les influences de la température, des pluies, de l'ensoleillement pendant la période de végétation peuvent être décelées. Mais l'analyse est délicate car il faut diviser la croissance des plantes en périodes et aussi étudier l'effet de facteurs climatiques combinés, hiver doux ou sec avec été chaud et pluvieux par exemple.

D'autre part, comme on l'a vu au paragraphe consacré à la biologie, la sélection des semences, l'étude des variétés, la mesure des croissances, l'examen combiné des facteurs végétal/sol/eau/nutrition ont transformé la recherche agronomique. Les champs d'expérience et les fermes pilotes n'auraient pas été possibles sans l'outil statistique. Et un des plus grands statisticiens anglais, R. A. Fisher, a consacré une grande partie de sa carrière à la direction du Centre d'Expérimentation Agricole de Rothamsted.

e) CONTROLE DES FABRICATIONS. — Depuis 1931, date du premier ouvrage sur la question (W. Shewart, *Economic control of the quality of manufactured products*), les travaux relatifs au contrôle des fabrications se sont généralisés avec une extraordinaire rapidité.

Certes, de tous temps, l'homme s'est préoccupé de la qualité

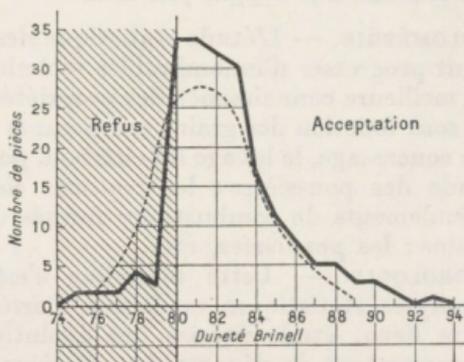


FIG. 5. — Contrôle : refus ou acceptation de pièces en cuivre selon leur dureté

de ses produits, mais il n'y a guère que 30 ans qu'il applique des méthodes statistiques compatibles avec la production en série. On ne peut pas vérifier, par exemple, tous les boulons qui sortent d'un atelier : on procède à des contrôles sur échantillons.

La production d'une machine ou d'un atelier se conforme à certaines normes ; mais sous l'effet de nombreuses petites causes accidentelles, la distribution d'une fabrication autour de cette norme donne une courbe gaussienne.

M. Swan donne un premier exemple de l'intérêt de cette distribution. Dans une usine anglaise, on faisait des essais de dureté Brinell sur certains objets en cuivre ; ceux dont la dureté était inférieure à 80 étaient mis au rebut. En étudiant la courbe des résultats, le service de contrôle remarqua une discontinuité aux alentours de la dureté 80. Il y avait eu « coup de pouce » de la part des responsables des ateliers qui voulaient faire passer au-delà du seuil de rejet les pièces légèrement inférieures à la norme. Après contrôle, il apparut que les pièces se distribuaient bien selon une loi normale (courbe en pointillé sur la fig. 5).

Si la distribution des produits fabriqués est assimilable à une distribution normale, cette distribution est entièrement définie par 2 indicateurs : celui de position (moyenne) et celui de dispersion (écart-type) (voir sur eux titre V, chap. I).

Le technicien fixe pour toute fabrication un intervalle de tolérance : par exemple telle pièce ne doit pas avoir $3/10$ de mm en plus ou en moins que la dimension fixée. Pour des raisons de rentabilité, on dira ensuite : les machines doivent être telles qu'il n'y aura que 3, 4 ou 5 % au maximum de pièces à rebuter. Sur la figure 6 on voit que la machine A, jugée aux essais, est assez précise : la dispersion de ses produits est entièrement contenue dans l'intervalle de tolérance. La machine B n'est pas assez précise ou encore elle est usée (une partie notable de ses produits est en dehors de l'intervalle). Quant à la machine C, on voit qu'elle est assez précise, mais sa fabrication, peu dispersée, doit être centrée : elle a été dérégulée.

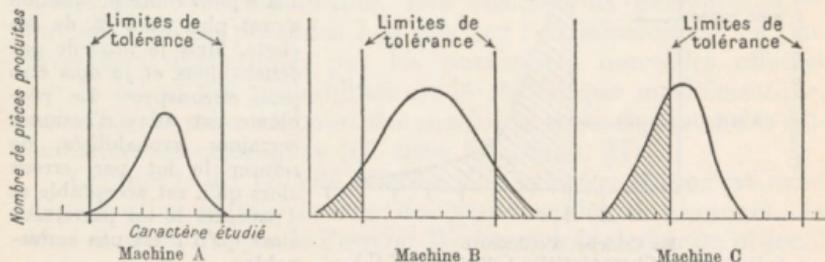
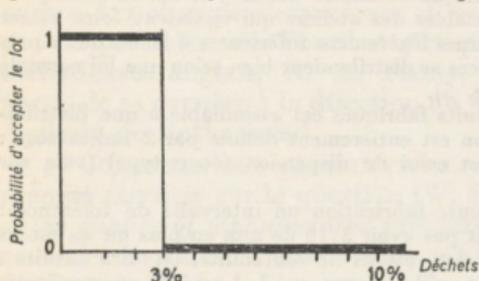


FIG. 6

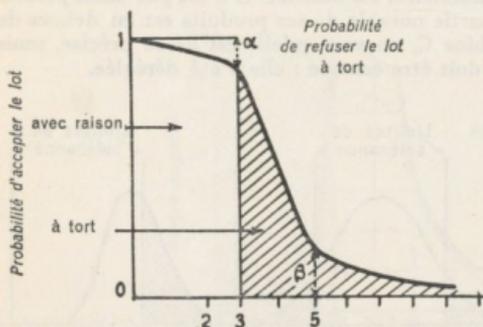
Il est bien évident qu'on ne peut contrôler toutes les pièces produites ; on prend seulement un lot ou un échantillon qui, s'il est bien constitué, donne, avec une forte probabilité qu'on se fixe à l'avance, la valeur vraie pour la population entière des pièces fabriquées. Tout entrepreneur arbitre entre le coût d'un échantillonnage lui donnant de grandes chances de précision et l'incertitude attachée à des échantillonnages moins coûteux.

Comme dans la plupart des démarches illustrées par les exemples précédents, la conduite de l'entrepreneur est de réaliser un test de signification.

Voici la présentation simplifiée d'un exemple de J. Mothes. Supposons qu'on décide de rejeter entièrement une production si elle contient plus



A) Jugement d'acceptation ou de refus sur échantillon unique (« Simple Sampling Plan », S.S.P.)



B) Courbe d'efficacité (« Operating Characteristic Curve », O.C.C.)

FIG. 7

de 3 % de déchets. La règle de test, pour être parfaite, doit conduire à la certitude (c'est-à-dire à la probabilité 1) d'accepter le lot quand il présente effectivement une proportion de déchets inférieure à 3 %, ainsi qu'à la certitude de refuser le lot (c'est-à-dire à la probabilité 0 de l'accepter) s'il en contient une proportion supérieure à 3 %. Si je connaissais exhaustivement le lot, je trancherais nettement comme sur le graphique 7 A : la probabilité d'accepter le lot passerait brutalement de 1 à 0 pour toute production ayant plus de 3 % de déchets. Mais je procède par échantillons et je dois être plus circonspect. Le problème est alors d'assumer certaines probabilités, de refuser le lot par erreur alors qu'il est acceptable et d'accepter le lot par erreur alors qu'il n'est pas convenable.

Supposons que mon

échantillon de 100 pièces m'indique 7 déchets : il y a encore une certaine probabilité pour qu'il soit tiré d'une population n'ayant que 3 % de déchets (un échantillon plus grand ayant cette même proportion de 7 % aurait moins de chances d'appartenir à cette population) : j'ai donc un certain risque de refuser le lot entier par erreur.

Mais un autre échantillon indiquant aussi 7 % de déchets peut aussi bien appartenir à une population contenant 10 % de déchets. J'ai donc un risque d'accepter le lot à tort. L'examen d'une *courbe d'efficacité* facilitera la décision. Le graphique 7 B montre une telle courbe qui remplace le choix brutal de 7 A ; elle sera la règle de jugement : avec l'échantillon tiré (7 %) « nous constatons que si notre lot présente 2 % de déchets (lot acceptable) nous avons une petite probabilité α de le refuser par erreur. Nous constatons également que si notre lot présente 5 % de déchets (lot défectueux) nous avons une petite probabilité β de l'accepter par erreur ». La courbe montre qu'on aurait pu choisir des probabilités α et β différentes, mais tout choix sera bon qui ne donne pas une importance aberrante à l'un ou à l'autre risque. De telles données suffisent à déterminer la règle du test à adopter : on sait donc, conclut J. Mothes, définir les règles de jugement à mettre en œuvre à partir de leur efficacité.

Il existe bien d'autres méthodes d'acceptation ou de refus et, bien entendu, la technique des tests s'applique, dans l'industrie, à beaucoup d'autres problèmes (contrôle des approvisionnements, des stocks, de la durée de vie du matériel, etc.).

§ 4. ACTIVITÉS DE GESTION ET D'ADMINISTRATION RECHERCHE OPÉRATIONNELLE

On pourrait rattacher les derniers exemples donnés à une discipline nouvelle qui pourrait être appelée la *science de la gestion et de l'organisation*. Sous l'appellation de *recherche opérationnelle*, ou préparation scientifique de la décision, on utilise de plus en plus la science du hasard pour élaborer la *décision* d'une autorité responsable. Des réalisations extraordinaires ont été rendues possibles à la fois par l'enrichissement de l'information statistique, par les possibilités nouvelles offertes par le calcul des probabilités et la statistique mathématique, enfin par la mise au point des machines statistiques et des calculatrices électroniques (cf. titre III, chap. II).

Comme le dit G. Pilé, *décider c'est arbitrer*. Mais c'est aussi dépasser les points de vue des responsables de secteurs. Le chef d'entreprise (ou le Premier Ministre) doit choisir un objectif unique, principal, qui n'est presque jamais analogue à ceux des services de la firme (ou des ministères) : le service approuvé

sionnement veut le maximum de stock de matières premières, le service fabrication le maximum de machines modernes, le service ventes, les produits souhaités par la clientèle à un prix bas. L'entrepreneur choisira *une* ligne de conduite, par exemple, dans une situation donnée, le plus bas prix de vente possible. Les exigences de stockage, d'équipement, etc., dont il faudra cependant tenir compte, en seconde urgence, apparaîtront alors comme des *contraintes* freinant ou limitant les possibilités d'atteindre l'objectif principal. Si en France la politique économique est si incertaine, c'est parce qu'il n'y a pas de mentalité recherche opérationnelle et que le gouvernement laisse émerger l'objectif d'un ministère — en règle générale celui qui est imposé par les Finances — ou quelquefois laisse poursuivre successivement les objectifs proposés par plusieurs (on investit, on distribue des subventions, on équilibre le budget, on augmente les impôts, on lutte contre la hausse des prix, etc.). S'il n'y a plus un chef d'entreprise pour vouloir à la fois maximiser ses ventes et ses achats et minimiser ses coûts et ses stocks, il y a encore des hommes politiques qui demandent plus à l'impôt pour demander moins au contribuable.

a) UN EXEMPLE DE GESTION DES STOCKS. — Dans une entreprise, le service commercial, effrayé de voir les clients commander certains jours quelques centaines d'articles, rien certains autres jours, demande au service fabrication un certain stock de sécurité pour ne pas manquer des ventes éventuelles. Mais ces stocks coûtent cher, les marchandises se détériorent et la direction veut les réduire au minimum. On examine donc le nombre d'articles vendus par jour, et en moyenne par commande. On dégage la tendance du marché (expansion par exemple). On étudie la distribution des ventes (fréquences des commandes/clients et des articles/jour). Là où le service commercial ne voyait aucune régularité, apparaît une distribution qui s'apparente à telle loi statistique (Gauss, ou Poisson par exemple). On peut donc calculer la probabilité de tel ou tel événement : un nombre n de commandes n'a que 5 chances sur 100 de se produire. Il y aura 5 jours par trimestre où la rupture de stock sera à redouter. Est-ce grave ? La perte sur les ventes est-elle alors supérieure aux coûts qu'impliqueraient des stocks plus importants ? Si oui, le stock optimum sera fixé à un niveau un peu supérieur.

b) THÉORIE DES JEUX. — Comme le dit G. Guilbaud, il n'y a pas que la complexité des problèmes et la présence de phénomènes aléatoires qui rend délicate la décision : « Pire que

la complexité, pire que les aléas, il y a les *adversaires* capables d'utiliser les mêmes techniques que vous et de réagir scientifiquement à vos décisions ».

Depuis le livre d'O. Morgenstern et J. von Neumann, *Theory of Games and Economic Behaviour* (1944), des applications nombreuses ont été tentées, qui toutes impliquent un recensement des politiques possibles et des réactions possibles de l'adversaire, une évaluation des conséquences possibles de chaque politique sur soi-même, une appréciation de la probabilité de voir l'adversaire faire tel ou tel choix, un couplage par paires de chaque politique propre/réaction adverse, avec affectation de coefficients « valeur de jeu ».

§ 5. DE L'ART DE LA GUERRE AUX BELLES-LETTRES ET A LA PÉDAGOGIE

a) GUERRE. — Bon nombre des techniques citées au paragraphe précédent sont nées pendant la seconde guerre mondiale et ont eu des buts stratégiques. Des spécialistes qualifiés ont pu écrire que la victoire des Alliés en 1945 a été due, en dépit d'une technique proprement militaire sans doute inférieure à celle des Allemands, à l'emploi de procédés de statistique mathématique et à la création de bureaux de recherche opérationnelle qui ont mis le hasard du côté des Alliés : lutte anti-sous-marine, organisation des convois, organisation optimum des escadres de bombardiers ont trouvé, grâce aux statisticiens, une efficacité accrue. Ajoutons encore que la préparation et la conduite de la bataille de Normandie ont été facilitées par l'emploi de calculatrices électroniques dans les bureaux de R.O. de l'Etat-Major.

b) LEXICOLOGIE ET MUSICOLOGIE. — Depuis longtemps, les philologues ont vu l'intérêt de dénombrements et calculs sur les textes. G. Raynaud de Lage, par l'étude des fréquences de mots, de rimes et de constructions, tend à montrer que le *Tristan* n'est pas tout entier l'œuvre de Béroul. Les spécialistes ont montré que les signes linguistiques (sons, mots, constructions grammaticales) se retrouvent avec une fréquence fixe dans un langage donné ; ces fréquences peuvent prendre des expressions mathématiques rigoureuses. Et lorsque, chez un auteur, on constate des écarts, il faut faire intervenir la statistique pour savoir si ces écarts sont ou non aléatoires. Il y a, par exemple, chez Racine une transformation profonde du langage

entre *Andromaque* et *Esther* — et dans la période qui sépare ces deux pièces, la date donnée à *Iphigénie* ne semble pas satisfaisante. P. Guiraud et R. L. Wagner posent la question de savoir si *Iphigénie* n'a pas été écrite avant *Bérénice*. Les mêmes spécialistes, examinant la richesse du vocabulaire chez Baudelaire et chez Mallarmé, notent une égale richesse ; les écarts sont dans la limite des variations dues au hasard, bien que de dispersions très différentes, Baudelaire tendant à répéter beaucoup plus souvent les mots usuels.

La structure statistique des langues, qui est à ses débuts, aidera non seulement à connaître et à situer des auteurs, à attribuer des œuvres de paternité incertaine, mais aussi à pénétrer, par le phénomène social qu'elle représente, la culture des différentes époques.

Un effort symétrique est tenté en musicologie et les distributions des sons, des rythmes, des constructions fait approcher des « lois du style » de chaque compositeur. Si ces lois étaient parfaitement accessibles on comprendrait la boutade du statisticien qui rêve de donner ces lois en programme aux calculatrices pour en tirer toutes les œuvres que J. S. Bach n'a pas écrites mais qui étaient « possibles »...

c) **DOCIMOLOGIE.** — Cette nouvelle discipline pose une question qui aurait paru bien étrange il y a quelques années encore : dans quelle mesure peut-on se fier aux examens et aux concours pour situer le niveau d'un élève ? La réponse est que les notes étant des *mesures*, elles doivent être traitées comme des populations de mesures, dans lesquelles il faut déceler d'une part les erreurs systématiques et d'autre part les erreurs fortuites.

La foi ancienne en la valeur des examens comme moyen de sélection venait de ce qu'on les assimilait à des opérations de *pesée* au moyen d'une balance précise et indéterminable pendant la durée totale des opérations. On sait aujourd'hui que cette balance n'a pas cette qualité, surtout lorsqu'elle est constituée par un groupe de correcteurs : le classement global dépend ainsi non seulement des candidats mais aussi des correcteurs.

Le problème est de soumettre la production des corrections à un véritable contrôle des fabrications, la note éliminatoire étant assimilable à la limite de tolérance du graphique 6.

D'innombrables méthodes sont possibles pour redresser les jugements individuels des professeurs et faire la péréquation entre les notes qu'ils ont données. Voici quelques exemples de disparités des « notes brutes » : M. Piobetta constate que dans « un des concours universitaires les plus élevés », avec double correction des copies, l'examineur A obtenait une moyenne arithmétique de 7,9 et l'examineur B une moyenne de 9,1.

D'autre part, sur 166 copies, 20 seulement avaient reçu la même note chez A et chez B. Il y avait 29 écarts de 1 point entre les 2 correcteurs, 26 de 2 points, 33 de 3 points et 21 de 5 ou 6 points, le rang d'un candidat reçu 17^e par A devenant le 95^e (ajourné) par B, etc.

Le même problème se pose dans les administrations publiques et privées pour la notation du personnel, dont dépendent primes et avancement, et des textes officiels prévoient la péréquation.

Une solution parfaite étant impossible, il s'agit de rechercher une approximation raisonnable dans 3 directions :

- 1) ajustement des paramètres de position : on peut admettre que les n groupes d'élèves notés par des correcteurs différents sont comparables au point de vue valeur centrale du groupe (médiane ou moyenne, voir pp. 347 à 363) ;
- 2) ajustement des paramètres de dispersion : on peut admettre que l'étendue des notes (de la plus faible à la plus forte de chaque groupe) et que l'écart des notes entre elles doivent être sensiblement égaux ;
- 3) élimination de la dissymétrie des distributions ; on en discute : pour l'élimination joue l'argument de la sévérité ou de la bienveillance qui affectent autant la forme de la distribution que sa valeur moyenne. Contre l'élimination on invoque l'idée que s'il faut effacer ce qui tient à l'équation personnelle des correcteurs (ajustements 1 et 2) il faut leur faire confiance quant aux différenciations qu'ils ont voulu faire à l'intérieur d'un même groupe, en marquant volontairement dans le groupe la présence de très bons ou de mauvais candidats.

Mais ces opérations de péréquation qui peuvent mobiliser toutes les ressources de la statistique probabiliste ne dispensent pas d'une précaution initiale : former les correcteurs en leur donnant, avant les épreuves, une courbe « idéale » des notes (centrée par exemple sur la note 12, avec des dispersions de 6 à 17) et en leur demandant de s'en éloigner le moins possible et seulement pour des motifs précis.

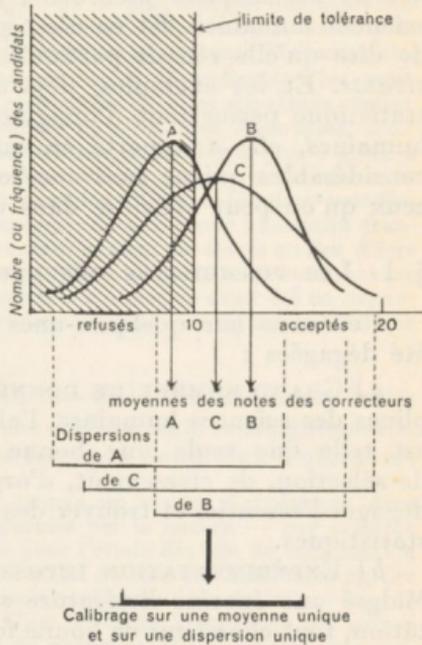


FIG. 8 — Péréquation des notes

SECTION III. — La Statistique dans les sciences sociales et en économie

L'expansion statistique dans les domaines les plus divers, décrite au chapitre précédent, s'est effectuée aussi dans les sciences humaines. Et ce n'est plus un paradoxe aujourd'hui de dire qu'elle réussit particulièrement là où il s'agit d'*êtres vivants*. Et les exemples, déjà donnés, des raisons du succès statistique permettent d'imaginer pourquoi, dans les sciences humaines, elle a apporté en quelques décades des résultats considérables qui ne sont encore que peu de chose à côté de ceux qu'on peut attendre dans un proche avenir.

§ 1. LES POSSIBILITÉS OFFERTES PAR L'OUTIL STATISTIQUE

Revenons sur quelques-unes des caractéristiques qui ont été dégagées :

a) GRAND NOMBRE DE DONNÉES. — Dans plusieurs disciplines des sciences humaines, l'abondance des faits enregistrés est telle que seule une bonne technique d'enregistrement, de sélection, de classement, d'organisation, de réduction peut dégager l'essentiel et trouver des régularités sous forme de lois statistiques.

b) EXPÉRIMENTATION IMPOSSIBLE OU TRÈS DIFFICILE. — Malgré une énorme littérature sur ce thème de l'expérimentation, tout chercheur de bonne foi doit convenir qu'en sciences humaines des « expériences » sont impossibles ou très rares. Des expressions telles que « psychologie expérimentale » ou « économie positive » abusent le public, si elles veulent indiquer autre chose qu'une réaction contre les méthodes purement abstraites.

P. Fraisse reconnaît qu'en psychologie « beaucoup de travaux que l'on qualifie d'expérimentaux ne sont, en réalité, que des *observations systématiques* », et D. Lagache affirme très justement que bien des conduites et situations « ne peuvent que difficilement ou pas du tout être *créées et contrôlées* artificiellement pour des raisons d'ordre moral ou technique : la psychologie de la jalousie amoureuse, du crime passionnel, du suicide a peu à attendre de l'expérimentation ».

Il en est de même en économie où les « expériences » ne sont que l'obser-

vation de phénomènes saisis dans l'enchevêtrement complexe du réel. « L'expérience monétaire française de décembre 1958 » n'est expérience que pour le responsable de la politique économique et pas pour le savant. Car dans le succès constaté, bien des facteurs étrangers ont joué à côté des facteurs créés (dévaluation et équilibre budgétaire) : dans une véritable expérience, il aurait fallu réaliser le « toutes autres choses restant égales », c'est-à-dire demander aux prix étrangers de ne plus bouger, aux individus de ne pas changer leur comportement, à la démographie de ne pas fournir plus d'hommes, au Marché Commun de ne pas modifier les courants d'échange, etc.

Or la statistique est bien l'outil qui permet de désenchevêtrer les paramètres et de jauger leurs influences respectives dans les situations réelles. La statistique replace donc l'observateur de situations *historiques*, donc uniques, dans une position *quasi expérimentale*. Mais il y a plus : grâce à l'emploi de modèles, le statisticien et l'économiste peuvent confronter avec cette *histoire* qui s'est effectivement déroulée d'autres *histoires possibles* qu'on aurait pu connaître *si*, avec l'exemple de la réforme monétaire française de 1958, les droits de douane européens avaient été élevés au lieu d'être abaissés, ou si les prix des matières premières avaient augmenté comme pendant la guerre de Corée, ou si l'économie américaine avait été en expansion de 1958 à 1961. A la *quasi-expérience* s'ajoute un immense domaine *para-expérimental*.

c) PETIT NOMBRE DE DONNÉES. — Si la statistique est indispensable là où le chercheur dispose d'abondants renseignements, elle l'est aussi là où il n'a que des faits peu nombreux à observer.

C'est le cas du psychologue qui trouve un ersatz d'expérience dans l'observation de situations limites, fournies par la nature — par exemple enfants élevés par des bêtes sauvages pour l'étude des lois de socialisation de l'individu, blessures à la tête pour l'étude du rôle des différentes parties du cerveau, infirmité des aveugles pour l'analyse du rôle de la vue dans le développement humain (P. Fraisse). C'est le cas de l'hygiéniste qui trouve dans la période de sous-alimentation de guerre et d'après-guerre les possibilités d'observation dans les écoles parisiennes pour en dégager la protection alimentaire des écoliers (J. Sutter). C'est le cas de l'économiste qui, du rapprochement des inflations galopantes en Allemagne (1923), en Hongrie, en Chine, en Grèce (après 1945), tire des explications de valeur générale.

Si à l'échelle macroscopique, économistes, sociologues et psychologues sociaux ne peuvent qu'observer certaines variables mal isolées du contexte général, et doivent donc demander à la statistique les procédés permettant de les séparer autant que possible, à l'échelle microscopique ils trouvent sur des *petits groupes* des conditions plus proches de celles du laboratoire, mais alors ils doivent demander à la statistique les subtiles

Participant d'une démarche de transmission de fictions ou de savoirs rendus difficiles d'accès par le temps, cette édition numérique redonne vie à une œuvre existant jusqu'alors uniquement sur un support imprimé, conformément à la loi n° 2012-287 du 1^{er} mars 2012 relative à l'exploitation des Livres Indisponibles du XX^e siècle.

Cette édition numérique a été réalisée à partir d'un support physique parfois ancien conservé au sein des collections de la Bibliothèque nationale de France, notamment au titre du dépôt légal. Elle peut donc reproduire, au-delà du texte lui-même, des éléments propres à l'exemplaire qui a servi à la numérisation.

Cette édition numérique a été fabriquée par la société FeniXX au format PDF.

La couverture reproduit celle du livre original conservé au sein des collections de la Bibliothèque nationale de France, notamment au titre du dépôt légal.

*

La société FeniXX diffuse cette édition numérique en accord avec l'éditeur du livre original, qui dispose d'une licence exclusive confiée par la Sofia – Société Française des Intérêts des Auteurs de l'Écrit – dans le cadre de la loi n° 2012-287 du 1^{er} mars 2012.

Avec le soutien du

