

# Introduction

## Des statistiques à la statistique

### Statistiques et statistique

***Les statistiques** sont des dénombrements de sujets, d'objets, d'événements dans une population ou des sous-populations. **La statistique** est une démarche permettant de recueillir, de traiter et d'interpréter les données qu'on recueille dans divers domaines où celles-ci présentent une caractéristique essentielle : la variabilité. Ainsi en est-il par exemple dans le domaine de la santé, où les individus diffèrent tous les uns des autres. Aboutir à une conclusion dans le domaine de la science ou de la décision à partir de cas tous particuliers est une gageure qui nécessite bien une démarche spéciale. C'est là le rôle de la statistique. Il en est de même notamment en agronomie, dans l'industrie ou l'économie.*

*La statistique peut certes porter sur des données récoltées par des statistiques, et celles-ci peuvent prêter à l'utilisation de la statistique, mais les deux démarches sont essentiellement différentes.*

*Les statistiques existent depuis des siècles, la statistique par contre repose sur un mode de pensée original qui ne s'est réellement développé qu'à partir du XIX<sup>e</sup> siècle.*

Ce rapport s'appuie sur un état de la contribution française au progrès des sciences, un état de la mise en œuvre des connaissances scientifiques par la communauté française ainsi qu'un état de l'enseignement. Ces états détaillent en particulier l'adaptation de l'activité française au contexte européen et international et examinent les perspectives possibles. Tant en sciences fondamentales qu'en technologies, le rapport propose des objectifs et des mesures susceptibles de les atteindre.

Avec le développement des moyens de télécommunications et des moyens de calculs, **les sciences et techniques de l'information** modifient sensiblement l'ensemble des processus de prises de décisions dans tous les secteurs de la vie sociale et économique. De par l'impor-

tance de leurs enjeux, cet ensemble de sciences et de techniques doit être distingué et caractérisé.

Au cœur des traitements de l'information, **les statistiques** sont utilisées de manière systématique, explicitement ou non. Depuis plus de trois siècles, elles ont pénétré *la gestion publique* (l'origine étymologique du terme le lie d'ailleurs à l'État). Établir et utiliser des bases de données détaillées et pertinentes, notamment dans leurs caractéristiques géographiques, a servi à développer rationnellement par étapes successives et irréversibles l'implantation des administrations, la construction des grandes infrastructures, les services d'éducation ou de santé, le contrôle des épidémies ou de la criminalité, etc. Dès lors que l'on a pris l'habitude de les rendre disponibles sous forme anonyme, ces « statistiques officielles » ont été employées aussi pour *la gestion privée* (études des marchés, implantation des établissements, calcul des risques, gestion financière, etc.). Simultanément, un ensemble de concepts, de procédures pour le traitement des données, de modèles et de méthodes pour leur utilisation apparaissait et se diffusait (indicateurs démographiques ou sanitaires, indices économiques, modèles de projection, calcul actuariel, modèles d'évaluation des prix des actifs, etc.).

Entraîné par le mouvement de démocratisation et par les demandes des chercheurs, notamment en sciences sociales, le pouvoir politique accepta progressivement que les données, collectées d'abord pour des besoins officiels, soient diffusées régulièrement pour *l'information du grand public*, et cela dans des conditions assurant de mieux en mieux leur objectivité et leur degré de signification. Les progrès considérables de cette information, qui ne sont évidemment pas tous dus aux statistiques, révèlent aujourd'hui un retard du public qui ne sait pas toujours en tirer bien parti, en raison parfois du retard culturel des informateurs intermédiaires. Ainsi en est-il des risques encourus dans la vie de tous les jours, par exemple dans les transports, dans la consommation médicale ou alimentaire, ou encore du fait de la qualité de l'air. Des risques réels sont souvent longtemps ignorés. Puis la diffusion de données les concernant, mais sans les paramètres statistiques adéquats pour l'évaluation de leur signification et pour celle des conséquences qu'elles pourraient annoncer, crée des réactions souvent excessives et irraisonnées. Toute une éducation de la lecture des statistiques et du traitement des risques doit être développée.

Dans de nombreux domaines de *la recherche scientifique et technique*, le recours aux statistiques s'est imposé au cours de ce siècle. Non seulement les statistiques économiques et sociales constituent la source principale d'observation des phénomènes globaux analysés par la science économique, la démographie ou la sociologie. Mais dans les sciences de la nature et de la vie également, les progrès de la méthode expérimentale passent souvent par la collecte rigoureuse et l'exploita-

tion adéquate de données nombreuses. Qu'il s'agisse d'agronomie, de biologie, de recherche médicale, de météorologie, de géologie, etc., l'efficacité de la recherche dépend de la mise en place de bons plans d'expérience et de l'application de bonnes procédures d'estimation ou de test. **La statistique mathématique** développe la recherche de tels plans et de telles procédures pour des contextes expérimentaux et disciplinaires très variés. La recherche technique bénéficie de progrès de même nature, par exemple pour les contrôles de qualité ou la mise au point d'automatismes fiables.

Constituer une communauté capable de mettre en œuvre correctement ces techniques et d'interpréter correctement les données, disposer d'une communauté d'experts pouvant être consultés pour le développement de nouveaux traitements, constituent des exigences pour une société moderne. La formation universitaire des statisticiens professionnels, des chercheurs de chaque discipline et des décideurs des divers secteurs d'activité, comme la formation des citoyens à la statistique, représentent ainsi des enjeux de première importance pour l'enseignement secondaire et supérieur en France.

Au cours de ce siècle, la méthodologie statistique s'est constituée en discipline à part entière. En parler comme d'une des branches des mathématiques appliquées, ainsi que cela est trop souvent fait en France, conduit à ignorer l'importance des acquis spécifiques non mathématiques, dont tout spécialiste des méthodes statistiques doit disposer pour diffuser l'emploi de ces méthodes, pour choisir entre elles à bon escient, pour en étudier les propriétés dans des contextes nouveaux, pour les faire progresser, pour en imaginer de nouvelles. Bien que ces compétences non mathématiques soient quelque peu spécifiques aux divers domaines de l'investigation statistique, elles ont assez de points communs pour s'épanouir au mieux dans le milieu des chercheurs en méthodologie statistique. Or si la communauté française des mathématiciens est classée aux tous premiers plans mondiaux, il nous reste beaucoup à faire pour parvenir à un jugement comparable quant à *la communauté des statisticiens méthodologues*.

Afin de procéder à une évaluation prospective de l'activité scientifique et universitaire française, à la fois sur un plan académique et sur le plan plus général des besoins de la société française, le groupe thématique « statistique » s'est proposé de répondre aux questions suivantes :

- Qu'appelle-t-on statistique ? Quelles sont la nature et la qualité de la recherche en statistique en France ? Quelle est sa place en Europe et dans le monde ?
- Quelles sont, en France, la nature et la qualité de la mise en œuvre des méthodes statistiques, dans quelques grands secteurs de l'économie et de la vie sociale ?

– Quelles sont, en France, la nature et la qualité de la mise en œuvre des méthodes statistiques dans la recherche scientifique et technique ?

– Quelles sont, en France, la nature et la qualité de la formation initiale de l'enseignement des statistiques, du primaire au supérieur, et de la formation continue ?