

PRÉSENTATION DU NUMÉRO

Musique Maestro !

Avec ce septième numéro, le *Courrier des statistiques* fête son troisième anniversaire. L'ambition de la revue est toujours d'aborder un large panel des grandes problématiques auxquelles se confronte la *statistique publique*. Elle propose une tonalité pédagogique, ouverte à la variété des sujets, des auteurs, des points de vue. Elle s'adresse au statisticien, débutant comme expert, et au citoyen, même si ce dernier pourra parfois trouver un peu aride sa lecture. Elle permet de témoigner de notre capacité collective à évoluer et innover tant sur les méthodes et outils, que sur des questions institutionnelles ou encore juridiques. La revue veille à rester attentive aux pratiques extérieures, en France comme à l'étranger, afin de se positionner vis-à-vis de notre communauté, de nourrir nos propres réflexions et de témoigner de nos travaux.

La revue, quoi de plus normal, s'intéresse particulièrement à la donnée et les différents moyens de la collecter ou de la produire. Elle traite, dans ce numéro, d'une évolution majeure de ces dernières années : l'intégration du multimode dans les enquêtes, à savoir la mise en musique, dans ces dernières, de modes complémentaires de collecte. Comme dans le numéro précédent, elle cherche aussi à savoir comment les statisticiens œuvrent de plus en plus pour tirer parti des gisements de données qui existent déjà mais qui restent insuffisamment valorisés. Enfin, on verra comment une grande opération statistique, le recensement agricole, se modernise.

La donnée est donc centrale dans ce numéro, elle qui constitue le cœur du métier du statisticien. Pour autant, ce dernier doit, aujourd'hui encore plus qu'hier, adopter une large gamme d'instruments. Au sein de son répertoire, la maîtrise des technologies de l'informatique dans les nuages (le *Cloud computing*) ou des développements informatiques les plus récents renforcent son autonomie, sa responsabilité et le champ des possibles avec lequel il pourra jouer et orchestrer ses traitements. La capacité à jouer de concert avec d'autres disciplines académiques est mise en évidence dans ce numéro, avec l'exemple de la psychométrie dans la mesure des compétences des élèves. Enfin, les statisticiens, issus d'organismes différents, doivent diffuser des données incontestables et cohérentes. Pour cela ils développent et adoptent un solfège commun pour ranger, classer et analyser les données : ce solfège, c'est la nomenclature. Le numéro N7 l'illustre avec la nouvelle nomenclature statistique des infractions, désormais commune à tous les acteurs de la statistique pénale.

En prélude de ce numéro, deux articles traitent de l'introduction d'internet et singulièrement des protocoles multimodes dans la collecte des enquêtes. **François Beck, Laura Castell, Stéphane Legleye et Amandine Schreiber** procèdent à une large revue de cette évolution : le multimode, mélange de collecte en face-à-face, par téléphone, papier ou internet, répond au contexte de difficulté croissante à contacter les ménages. Mais par ailleurs, il complexifie le processus tout entier, tant dans la définition du protocole de collecte que dans les traitements statistiques. **Éric Sigaud et Benoît Werquin** détaillent ensuite la nécessaire mise au diapason des différentes étapes de la collecte, pour les équipes en charge de la concevoir et de la mettre en œuvre. Cela requiert *ex-ante* de conceptualiser chacune de ses phases, depuis la conception et génération automatique des questionnaires jusqu'à leur traitement et la consolidation des données. Ils nous offrent, ce faisant, une nouvelle interprétation de la démarche de pilotage par les métadonnées actives.

Le recensement agricole de 2020 est à l'honneur de ce numéro dans un article de **Hervé Le Grand**. Cette opération majeure de la statistique agricole est au cœur de son système d'information : elle en est en quelque sorte son métronome qui donne le rythme des autres enquêtes et garantit de produire des données cohérentes, au niveau national et européen. La dernière édition porte cinq innovations majeures, se répercutant sur les enquêtés, les enquêteurs et les statisticiens. Les données ont été collectées, pour la première fois, majoritairement par internet ou par téléphone. À l'heure où nous écrivons ces lignes, une sixième innovation, la *data visualisation*, apporte une touche finale à ce recensement décidément riche en évolutions.

Un trio de pirates, **Frédéric Comte, Arnaud Degorre et Romain Lesur**, nous emmène voyager dans le SSPCloud. Environnement informatique d'aide à l'expérimentation sur les nouvelles méthodes de la *data science*, le SSPCloud est constitué d'un ensemble de ressources informatiques permettant de réaliser des prototypes, de tester des traitements statistiques et de s'approprier de nouvelles pratiques de travail. Avec le SSPCloud, le statisticien s'inscrit dans un courant d'inspiration de type FabLab lui permettant de valoriser les nouveaux gisements de données. Ici on compose à plusieurs et l'adoption de solutions *open source* garantit les possibilités de réutilisation. Le SSPCloud, c'est enfin un mélange fertile des deux univers professionnels : celui de la statistique et celui de l'informatique.

Dans la droite ligne de l'article sur le SSPCloud, **Emmanuel L'Hour, Ronan Le Saout et Benoît Rouppert** nous parlent du statisticien selfeur. Le métier de statisticien requiert aujourd'hui une bonne maîtrise des outils informatiques. Terminé le temps d'une interprétation *a cappella* de sa partition. Il doit coder selon les règles de l'art car si les programmes informatiques écrits doivent permettre de produire des résultats, ils sont, au-delà des livrables, des éléments de preuve de la qualité des traitements et doivent pouvoir être réutilisés pour d'autres travaux ou réinterprétés par d'autres statisticiens selfeurs.

Le sixième article de cette revue nous conduit dans l'univers de l'exploitation et de l'appariement de sources administratives exhaustives. **Mathias André et Olivier Meslin** décrivent le projet qu'ils conduisent en duo : créer une nouvelle base statistique, afin de pouvoir étudier le patrimoine immobilier des ménages et le profil redistributif de la taxe foncière. Attachés à mobiliser les sources administratives disponibles, ils ont fait l'expérience des obstacles rencontrés avant d'accéder aux données ainsi que des chausse-trappes de l'appariement et du traitement statistique de fichiers venant d'univers différents, conçus pour d'autres usages. L'article décrit avec précision les étapes de l'épopée : il retient les bonnes pratiques qui leur ont permis d'aboutir à une base de production, laquelle vient désormais compléter le panorama de l'information statistique concernant le patrimoine des ménages. Il met également en évidence des enseignements de ce projet à destination du statisticien souhaitant mener des travaux sur des bases administratives.

Les travaux du statisticien prennent parfois un chemin singulier, du fait que l'objet que l'on souhaite mesurer ne pré-existe pas à l'opération de mesure elle-même. **Thierry Rocher** décrit les solutions déployées par le service statistique de l'Éducation, pour approcher une mesure des compétences des élèves. Il ouvre une fenêtre sur les concepts relevant du champ de la psychométrie. Il nous décrit les choix opérés – procédures et modélisations spécifiques – pour arriver à produire des évaluations standardisées des compétences des élèves. Ce faisant, il nous rappelle l'étendue des dispositifs, nationaux et internationaux, qui cherchent à définir des statistiques comparables, et utiles à tous les niveaux, depuis celui de l'enseignant et du chef d'établissement, jusqu'à celui d'un ministre.

Les dernières mesures de ce numéro nous entraînent du côté d'un défi assez rare dans la vie d'un statisticien : celui de l'élaboration d'une nomenclature statistique. Jusqu'alors, ministère de l'Intérieur et ministère de la Justice utilisaient des nomenclatures de diffusion différentes, ce qui empêchait de disposer de statistiques fines cohérentes tout au long de la filière pénale. **Benjamin Camus** décrit comment l'ONU a mis au point en 2015 une nomenclature internationale, s'affranchissant des différences de législations pénales, en choisissant une approche fondée sur le comportement de l'auteur de l'infraction. Celle-ci a donné le *tempo* et fourni l'occasion de lancer le chantier en France : un groupe de travail interministériel en a défini une déclinaison française ancrée sur une codification détaillée du droit pénal. En décembre 2021, la Nomenclature française des infractions a vu le jour : articulée avec la nomenclature internationale pour les grandes catégories, mais comprenant un niveau de détail plus pertinent dans le contexte français, elle porte en elle le germe de statistiques réconciliées.

Odile Rascol
Rédactrice en chef, Insee
