

## La statistique dans la cité n° 30 - juin 2022

*Lettre bimestrielle du groupe « Statistique et enjeux publics »*

### Sommaire du n°30 – juin 2022

#### Éditorial

#### Vie des institutions :

- *Un message de nos collègues du Service Statistique d'Etat de l'Ukraine*
- *Vers un espace européen des données de santé ?*

#### Méthodes :

- *Les prix des billets de train augmentent-ils ?*

#### Outils :

- *Mesurer les discriminations : un défi statistique et méthodologique que tente de relever un groupe de travail du CNIS*
- *Mesurer la mobilité sociale entre générations : l'Insee utilise l'Échantillon Démographique Permanent enrichi par des données fiscales*
- *Combien de décès sont dus à la pandémie de Covid-19 ?*

#### Humour ?

- *Quatorze personnes noyées dans la Chicago River à cause du recensement américain de la population de 2020 ?*

#### Nous avons lu :

- *Hommages rendus à deux éminents statisticiens qui nous ont quitté récemment, Jean-Claude Deville et André Vanoli*
- *Un dossier dans le supplément Sciences et Médecine du quotidien Le Monde daté du 18 mai 2022 sur l'objet de recherche que constituent les modes de scrutin*
- *Le Monde, dans son édition datée du vendredi 3 juin 2022, fait état du Rapport sur les riches en France publié par l'Observatoire des inégalités*

#### Vie du groupe

**Tous les numéros de l'infolettre sont téléchargeables sur le site du groupe SEP ([cliquez ICI](#))**

### Éditorial



**Слава Україні**

La Statistique dans la Cité ne pouvait pas ignorer la situation dramatique que vit l'Ukraine depuis le 24 février, il y a bientôt quatre mois. Cette situation qui touche l'ensemble de la société ukrainienne affecte bien entendu nos collègues, statisticiens publics ukrainiens, dans leur vie professionnelle comme dans leur vie familiale. La rédaction de notre infolettre a donc décidé de leur ouvrir ses colonnes, ce qu'ils ont confraternellement accepté. Le premier des articles que vous trouverez dans ce numéro a été rédigé par un responsable de la division des relations internationales et de l'intégration européenne du Service Statistique d'Etat de l'Ukraine (SSSU) à la demande de la rédaction de *La Statistique dans la Cité*. Dans cet article, nos collègues ukrainiens font part des difficultés de collecter et de traiter les données mais aussi de leur volonté de continuer à proposer une information aussi fiable et précise que possible aux décideurs et au public le plus large, malgré les difficultés liées à la guerre. Nos collègues font aussi appel à la solidarité et à l'assistance de leurs collègues européens.

Une délégation ukrainienne a pu participer à la 18e Conférence biennale de l'*International Association for Official Statistics* (IAOS), association de la famille de l'*International Statistical Institute*, qui avait lieu

à Cracovie du 26 au 28 avril derniers. Une des sessions de cette conférence avait précisément pour thème *Statistics and Birth of a Nation* : de nombreuses nations soudainement apparues ou réapparues au fil de l'histoire ont dû affirmer aux yeux de la communauté internationale que leur existence reposait sur des bases solides ; les statisticiens ont alors joué un rôle important, la production d'un ensemble cohérent et complet de statistiques étant clairement pour ces nouvelles nations la preuve de leur existence. Cette session s'appuyait sur les exemples de la création de la Belgique en 1830, de la renaissance de la Pologne en 1918, sur la mise en place d'une Autorité palestinienne en 1994 et sur la construction européenne. L'Ukraine aurait pu trouver sa place dans ce panel.

On trouvera aussi dans ce numéro de *La Statistique dans la Cité* un article présentant le projet européen d'un espace commun des données de santé, qui est d'une certaine façon une pièce à apporter au feuilleton *Health Data Hub* de notre infolettre. L'article suivant relate la controverse entre l'Insee et la SNCF sur le prix des billets de train (ont-ils ou non augmenté ?) ; il illustre bien la difficulté à calculer des indices synthétiques. Suivent trois articles dans la rubrique « Outils » : mesurer les discriminations, mesurer la mobilité sociale entre générations, compter les décès dus à la pandémie de la Covid-19. Nous terminons par un billet d'humour, des notes de lecture et des informations sur la vie du groupe *Statistique et Enjeux publics*, éditeur de cette infolettre.

*La Statistique dans la Cité* va prendre maintenant ses congés d'été et nous vous retrouverons à la mi-octobre avec notre n° 31. Et comme d'habitude, nous souhaitons que nos lecteurs puissent nous faire part de leurs réactions et de leurs commentaires sur nos activités et en particulier sur cette infolettre en nous écrivant à l'adresse [sep@sfds.asso.fr](mailto:sep@sfds.asso.fr)

**Pour nous écrire :** [sep@sfds.asso.fr](mailto:sep@sfds.asso.fr)

## **Vie des institutions**

### **Un message de nos collègues du Service Statistique d'Etat de l'Ukraine**

*Cet article a été rédigé par un responsable de la division des relations internationales et de l'intégration européenne du Service Statistique d'Etat de l'Ukraine (SSSU) à la demande de la rédaction de La Statistique dans la Cité. Il a été rédigé en ukrainien et traduit en anglais par les soins du SSSU. Nous avons choisi de publier le texte anglais tel qu'il nous a été envoyé, mais nos lecteurs peuvent, s'ils le souhaitent, accéder à une traduction en français établie par la rédaction de l'infolettre [en cliquant ici](#).*

Despite the war started by Russia in 2014 and its full-scale invasion of Ukraine that has been going on for more than three months, Ukrainian statisticians have not stopped working as society, country's leadership and our international partners need statistical information. The state statistics bodies of Ukraine have faced new difficult challenges during the war.

First of all, we had to save people's lives. At present, many regional offices of the State Statistics Service are unable to fully operate due to the threat to the lives and health of employees because of the constant shelling and bombings by « ruscists <sup>(1)</sup> ». As a result of the shelling in late May, the building of one of the territorial statistical offices was damaged. Fortunately, none of the employees were injured.

To ensure the process of statistical information production, the State Statistics Service introduced a support and assistance scheme. Under this scheme, another territorial office beyond the area of active hostilities was assigned to assist a territorial office that was not able to perform the task of collecting and processing information due to enemy shelling. This method allowed ensuring the continuity of the production of statistical information across Ukraine.

We lost access to traditional data sources due to the war. For instance, the submission of statistical reports by respondents, during Martial Law in Ukraine, has been postponed at the legislative level for the period of up to three months after termination or cancellation of Martial Law or the State of War. In this respect, sample surveys of the population (households) as well as the calculation of demographic indicators have temporarily been suspended in Ukraine. The results of these surveys are the basis for analyzing the socio-economic situation of the country, various aspects of the life of the population and conducting international comparisons (e.g. employment and unemployment, household living conditions, poverty and material deprivation, etc.).

At the same time, prices are constantly registered (this task is carried out by specialists of the State Statistics Service by visiting outlets or via the Internet), which allows the monthly publication of the

## Consumer Price Index.

As a result of Russia's military aggression, active migration processes are taking place in the country caused by significant forced internal and external migration of the population; in some regions, cities and villages have been completely destroyed, which makes it impossible to conduct sample surveys of the population (households) and to make observations in the context of demographic statistics as well as calculations of the population, births and deaths, etc.

In addition, it is crucial for Ukraine to assess the losses of the country's economic capacity caused by the devastation or shutdown of enterprises as well as the destruction of infrastructure. According to expert estimates, the total losses of Ukraine's economy due to the war vary \$564 billion to \$600 billion, as of May 25.

In this context, it is very important to have relevant statistical information. Therefore, we are now urgently reviewing the methodological base, looking for other ways and methods of data collection and calculation, new alternative sources of information in combination with cutting-edge technologies.

Thus, the methodological provisions for conducting the state statistical observations have been supplemented with a new section on the special aspects to conduct surveys under emergency and force majeure (in particular, under Martial Law), which determines the procedure in the total or partial absence of data for statistical production. We will be grateful to your readers – all our colleagues for methodological, expert and other assistance in that endeavor.

The State Statistics Service of Ukraine has received humanitarian and technical assistance from our friends and colleagues from the statistical offices of Poland, Lithuania, Norway and the UK. We sincerely thank everyone and are confident that together we will overcome the aggressor and bring the long-awaited victory closer. We continue to work.

Glory to Ukraine! Слава Україні

*(1) Mot-valise très utilisé en Ukraine depuis le 24 février. Ce mot a été composé à partir de « Russia » et de « facist » [Note de la rédaction].*

### **Vers un espace européen des données de santé ?**

La Commission européenne a annoncé le 3 mai un projet de règlement à soumettre au Parlement et au Conseil européens en vue de créer un [Espace européen des données de santé \(EEDS\)](#).

Il s'agit pour les pays membres de l'Union Européenne (UE) d'unifier autant que possible :

- les règles de droit applicables aux données de santé, notamment les droits d'accès, d'opposition et de rectification ainsi que les modalités d'exercice de ces droits, dont [une étude](#) a montré la très grande diversité entre les États membres, malgré la référence commune au Règlement général sur la protection des données (RGPD) du 27 avril 2016 ;
- les règles de sécurité et de contrôle pour l'habilitation, l'identification et l'authentification des personnes voulant accéder aux données ;
- les aspects sémantiques (nomenclatures, formats) des enregistrements électroniques de données de santé afin que les logiciels concernés (à l'hôpital comme en ambulatoire...) deviennent interopérables et puissent être certifiés ;
- voire (dans l'option maximale, non retenue à ce stade par la Commission) de mettre en place une plateforme informatique commune pour les données transfrontalières (à la manière du Health Data Hub français)...

Le projet de la Commission vise toutes les données de santé : celles dont l'usage est individuel (les usages premiers des données de santé sont généralement les soins de santé et l'assurance maladie, au bénéfice de personnes identifiées) et celles dont l'usage est collectif (on réutilise fréquemment à cette fin des données individuelles « pseudonymisées »).

Les besoins sont aujourd'hui évidents et incontestables que ce soit pour se faire soigner dans un autre pays européen en donnant accès à son dossier médical, ou pour réutiliser des données de santé provenant de plusieurs pays, à des fins de recherche, d'évaluation, d'innovation ou de sécurité sanitaire.

Une première difficulté est que les usages individuel et collectif, s'ils s'appuient largement sur les mêmes données, diffèrent grandement quant aux règles de droit et de sécurité applicables. Parmi les usages individuels (donc nominatifs), on trouve notamment les dossiers médicaux, qu'il s'agisse de dossiers tenus localement (par mon médecin, mon hôpital...) ou de dossiers partagés entre un grand nombre de professionnels de santé ou médico-sociaux, comme aurait dû l'être, en France, le Dossier Médical Partagé, ou DMP <sup>(2)</sup>. Or le principe d'un dossier médical partagé est bien sûr que chacun puisse accéder en ligne à son propre dossier, mais aussi que tous les professionnels qui contribuent à

soigner le titulaire et ont besoin des ces informations puissent eux aussi y accéder facilement. Cela suppose au minimum que tous ces professionnels soient bien identifiés et s'authentifient en ligne correctement (disposant d'une trace *prouvant* que le Dr A s'est connecté tel jour au dossier de Mme B, un juge appréciera si cela répondait à un besoin légitime). Si ce système, que la France n'a pas réussi à mettre en place, est étendu à toute l'UE, cela signifie que des millions de personnes (tout professionnel de santé ou *toute personne pouvant techniquement se faire passer en ligne pour un professionnel de santé identifié*) pourraient accéder à mon dossier médical ! Que le dispositif soit techniquement au point et inspire confiance aux Européens (et que les problèmes de traduction soient aussi résolus) prendra... un certain temps. C'est déjà difficile à réaliser dans un pays comme la France ; ça le sera encore plus si on veut éviter que n'importe qui en Slovénie puisse accéder au dossier médical d'un citoyen du Portugal ou inversement... Cela étant, il n'est pas interdit de viser aussi des solutions assez simples pour démarrer et on peut espérer qu'on se trompera moins à vingt-sept...

Quoi qu'il en soit, pour tous les usages des données de santé, l'intérêt d'unifier les règles juridiques, les règles de sécurité et les règles sémantiques <sup>(3)</sup> semble incontestable même si cela prendra là aussi du temps. On peut espérer ainsi que le futur [système européen d'identification des personnes](#) simplifiera la gestion et le contrôle des accès aux données, et que la gouvernance commune trouvera une solution juridique satisfaisante pour les appariements entre bases de données de sources différentes (un sujet pour lequel en France un équilibre entre les préoccupations de la Cnil et celles des chercheurs en santé n'a pas encore été trouvé).

En revanche, il y aurait certainement une autre difficulté si l'UE voulait promouvoir une plateforme informatique commune pour gérer les appariements « transfrontaliers » (entre données de pays différents). Le principe de subsidiarité milite plutôt pour une saine concurrence entre les plateformes techniques privées ou publiques des différents pays ou résultant déjà d'une coopération entre pays membres dès lors que ces plateformes respectent les règles. A cet égard, l'expérience française du Health Data Hub qui n'a pas véritablement démarré en raison notamment des réticences des gestionnaires de données (hospitaliers et autres) montre plutôt que la solution centralisée d'un jardin à la française (s'appuyant en outre sur un prestataire étatsunien !) n'est pas la voie à suivre : la Commission écarte l'option à ce stade mais rien n'est joué.

*(2) Le DMP est en voie de remplacement par l'Espace numérique de santé (plus simple et plus complet mais dont il reste à voir s'il réussira).*

*(3) Par exemple dans presque tous les pays européens on code les diagnostics et les actes médicaux dans les hôpitaux pour identifier l'activité et allouer les budgets ou facturer les soins mais chaque pays utilise ses propres nomenclatures...*

## **Méthodes**

### **Les prix des billets de train augmentent-ils ?**

Oui, selon l'indice des prix de l'Insee : pour le poste de dépense « Transports de passagers par train », cet indice affiche une hausse de 6 % entre avril 2019 et avril 2022. Non, selon les dirigeants de la SNCF : son PDG Jean-Pierre Farandou a affirmé, en comparant 2019 et 2022, « les prix moyens ont objectivement baissé de 7 % ! » <sup>(4)</sup>. Qui croire des deux ?

En parlant de « prix moyens », J.P. Farandou nous met sur une piste. Chaque année, beaucoup de billets sont vendus, à des prix divers ; pour connaître l'évolution des prix entre deux années, le plus simple n'est-il pas de comparer les prix moyens des billets entre les deux années ? Ne faites pas cela, répondent les statisticiens ! Avec cette méthode, vous mélangez les évolutions de prix avec les changements des choix des consommateurs. Par exemple, si les voyageurs optent désormais pour des trajets plus courts, le prix moyen des billets vendus va baisser, même si les tarifs sont inchangés. La bonne méthode consiste à comparer au fil du temps des billets comparables, à calculer des indices d'évolution élémentaires pour chaque groupe de billets comparables, et finalement à regrouper ces indices élémentaires en un indice d'ensemble en tenant compte de l'importance respective des différents groupes de billets comparables. Tous les indices de prix officiels dans tous les pays procèdent ainsi.

Il est difficile de croire que la SNCF ignore ce piège bien connu affectant l'évolution des prix moyens. On peut supposer que la baisse « objective » de 7 % citée par son PDG s'entend « pour des trajets identiques » <sup>(5)</sup>. Même en ce cas, les billets sont-ils bien comparables ? C'est de nouveau un dirigeant de la SNCF qui nous donne à réfléchir. Christophe Fanichet, PDG de SNCF-Voyageurs, parlant cette fois de la comparaison entre avril 2021 et avril 2022, reconnaît une hausse, mais met en cause la crise du Covid : « Peu de trains circulaient l'an dernier, compte tenu de la situation sanitaire, d'où un nombre élevé de places disponibles et donc des prix moins élevés », explique-t-il <sup>(6)</sup>. L'argument est



curieux : peu de trains, d'où beaucoup de places disponibles ? Quoi qu'il en soit, le PDG de SNCF-Voyageurs attire l'attention sur une question essentielle : le prix s'adapte désormais à la demande. Depuis plusieurs années en effet, il n'y a pas un prix unique pour un trajet donné en TGV : le prix varie selon le taux de remplissage du train et le moment où le client prend sa réservation, plus ou moins tôt avant le départ du train ; ce que l'on appelle « l'antériorité ». La SNCF pratique le « yield management »<sup>(7)</sup>.

Les statisticiens ont tenu compte de cette nouveauté. Depuis le début de 2020, l'indice des prix des billets n'est plus calculé d'après le tarif publié par la SNCF, tarif qui ne contient que des prix de référence dont la réalité peut s'écarter ; il est calculé à partir d'un échantillonnage de prix réellement demandés. Chaque mois, un robot lance plus de 10 000 requêtes auprès de sites Internet de vente de billets de train. Cette collecte de données automatisée permet de constituer des échantillons de billets pour 250 trajets et quatre antériorités (2 jours, 10 jours, 30 jours et 60 jours avant le départ du train). Et les indices élémentaires sont calculés pour chaque trajet et chaque antériorité<sup>(8)</sup>. Ils sont ensuite agrégés comme il a été indiqué ci-dessus. L'indice qui en résulte reflète l'évolution du prix effectivement demandé ; il ne peut pas être affecté par les éventuels changements du comportement d'anticipation des voyageurs.

Le « yield management » a pour effet que le prix est modifié très vite en cas de modification de la demande. Du coup, l'indice du prix des billets fluctue fortement d'un mois sur l'autre<sup>(9)</sup> : sa moyenne pour les six derniers mois disponibles s'établit à un niveau proche de celui de 2019, mais les prix selon les mois peuvent s'en écarter de plus de 5 %. La divergence avec l'estimation de la SNCF apparaît moins importante que si l'on utilise seulement le chiffre d'avril 2022 : néanmoins, elle reste considérable, et inexplicable.

(4) Relaté par « Le Figaro » le 23 mai 2022. Plusieurs autres médias ont fait état de la divergence : Libération, HuffPost, Le Journal du Dimanche, Le Canard Enchaîné...

(5) A notre connaissance, la SNCF n'a pas publié d'article précisant comment elle aboutit à ce chiffre.

(6) HuffPost, 29/5/2022, La SNCF répond aux critiques sur les prix des billets de train en forte hausse.

(7) Cette expression en anglais n'a pas de traduction française satisfaisante. L'expression « tarification différenciée » ne rend compte que d'une facette d'un mode de gestion de l'offre visant à maximiser le revenu de la société (« yield »).

(8) et en tenant compte aussi de la période (semaine ou week-end) et du type de voyageur (avec ou sans carte de réduction). Toutes ces informations ont été publiées en février 2020 par l'Insee : Informations Rapides n°39, 20/2/2020.

(9) Voir la série « Indice des prix, poste 07.3.1.1, IDBANK=001763668 » disponible sur le site de l'Insee.

## Outils

### **Mesurer les discriminations : un défi statistique et méthodologique que tente de relever un groupe de travail du CNIS**

Les discriminations se constatent dans de nombreux domaines : éducation, formation, emploi, logement, services publics ou privés, vie sociale, etc. Le Défenseur des droits a rédigé sur ce thème un rapport [Inégalités d'accès aux droits et discriminations en France](#), publié en 2019 par La documentation Française.

Dans le cadre de sa réflexion de moyen terme 2019-2023, le Cnis a souligné « la menace que constituent les discriminations pour la cohésion sociale » et souhaité « la construction d'un dispositif d'observation et de mesure des différentes formes de stéréotypes et discriminations ». Dans le prolongement des travaux menés par le Défenseur des droits, le Cnis a donc mis en place fin 2021 un [groupe de travail sur l'observation et la mesure des discriminations](#).

La [lettre de mission](#) fixant le mandat et les objectifs de ce groupe lui demande de préciser les diverses formes de discrimination les plus susceptibles de menacer la cohésion sociale, de dresser un état des lieux de la façon dont les dispositifs statistiques actuels sont aujourd'hui en mesure d'en rendre compte et identifier les manques, d'examiner de quelle manière ces discriminations sont mises en évidence dans les dispositifs d'observation de quelques autres pays, et enfin de formuler des recommandations pour améliorer l'observation et la mesure des discriminations. Ce groupe, qui a déjà tenu trois réunions, devra formuler des recommandations à l'ensemble des commissions du Cnis de l'automne 2022, et rendre son rapport final fin 2022.

Le groupe *Statistique et enjeux publics* prévoit de consacrer une séance des *Cafés de la statistique* à ce sujet important au cours du premier semestre 2023.

## **Mesurer la mobilité sociale entre générations : l'Insee utilise l'Échantillon Démographique Permanent enrichi par des données fiscales**

La reproduction des inégalités peut être appréhendée par la mobilité en matière de position dans l'échelle des revenus. C'est ce qu'a fait l'Insee à partir des données de l'[échantillon démographique permanent \(EDP\)](#) désormais enrichi par les revenus déclarés à l'administration fiscale. Il compare le « classement », dans l'échelle des revenus individuels, des enfants (autour de 28 ans) à celui de leurs parents <sup>(10)</sup>.

Cette [étude parue en mai 2022](#) met en évidence de nombreux mouvements dans l'échelle des revenus d'une génération à l'autre : « *Les enfants de familles aisées ont trois fois plus de chances d'être parmi les 20 % les plus aisés que ceux issus de familles modestes : les inégalités se reproduisent donc en partie d'une génération à l'autre. Cependant, pour un même niveau de revenu des parents, les revenus des enfants varient fortement. En 2018, parmi les jeunes issus des familles les 20 % les plus modestes, 12 % sont parmi les 20 % les plus aisés de leur classe d'âge.* » « *Toutes choses égales par ailleurs, la mobilité ascendante est d'autant plus forte que les parents ont des revenus du capital élevés, sont diplômés du supérieur, sont immigrés, ont été mobiles géographiquement, ou que les enfants résident en Île-de-France à leur majorité. À l'inverse, être une femme, avoir vécu dans une famille monoparentale, avoir des parents ouvriers ou employés, ou vivre dans les Hauts-de-France à sa majorité sont des facteurs qui réduisent les chances de s'élever dans l'échelle des revenus.* ».

[Sur son blog](#), l'Insee replace cette étude (dont les conclusions peuvent sembler contradictoires avec le sentiment de déterminisme social très fort en France) dans la littérature récente sur le sujet et fait le point sur ce qu'on sait de la mobilité intergénérationnelle des revenus en France.

*(10) Les données disponibles aujourd'hui ne permettent de mesurer directement la mobilité des enfants que sur les générations ayant moins de 30 ans, mais les données de l'EDP vont s'enrichir chaque année au fur et à mesure que ces générations vont vieillir.*

## **Combien de décès sont dus à la pandémie de Covid-19 ?**

Cette question, en apparence simple n'a pas de réponse évidente. Et il semblerait même que des organismes publics y apportent des réponses différentes. En fait, l'écart entre les chiffres fournis sur la période mars 2020 – décembre 2021 par l'Insee (95 000 décès supplémentaires par rapport à ce que l'on pouvait attendre sans épidémie) et Santé publique France et l'Ined (130 000 décès liés à la Covid-19) s'explique aisément, car les concepts retenus ne sont pas les mêmes.

L'Insee mesure le surplus de décès en 2020 et 2021 par rapport à ce qui aurait pu être constaté en l'absence de pandémie, en tenant seulement compte du vieillissement de la population, de l'évolution des quotients de mortalité par âge et du fait que 2020 est une année bissextile. C'est cette mesure qui conduit au chiffre de 95 000 décès supplémentaires sur la période. Santé publique France recense les décès attribués à la Covid-19 survenus à l'hôpital ou en Ehpad (124 000). Et l'Ined estime que 5 % des décès liés à la Covid-19 ont eu lieu à domicile, soit 6 000.

L'écart entre ces chiffres est lié à divers facteurs : certaines personnes fragiles auraient de toutes façons figuré numériquement dans les « décès attendus » et la pandémie n'a fait qu'anticiper de quelques semaines l'issue fatale ; des décès qui ne sont pas directement dus à la pandémie, en sont malgré tout la conséquence indirecte : report d'actes médicaux, baisse des dépistages de certaines maladies... ; à l'inverse le confinement dû à l'épidémie a permis d'éviter certains décès liés à d'autres causes : accidents de la circulation, maladies contagieuses. De plus, lorsqu'une personne atteinte d'une maladie grave contracte le virus dans ses derniers jours, il est parfois difficile de savoir s'il convient d'imputer le décès à la maladie préexistante ou au virus.

Les statistiques détaillées par cause de décès ne seront connues que dans plusieurs mois. Elles apporteront peut-être un nouvel éclairage sur l'impact réel de la Covid-19 sur la mortalité en France. D'ici là, on peut trouver quelques autres informations dans le [n° 1902 d'Insee-Première](#) paru le 19 mai 2022.

## **Humour ?**

### **Quatorze personnes noyées dans la Chicago River à cause du recensement américain de la population de 2020 ?**

Selon les résultats du recensement de la population de 2020 aux États-Unis, quatorze personnes habitent le « bloc <sup>(11)</sup> de recensement » 1002 situé dans le centre-ville de Chicago. Mais ce bloc n'est en réalité qu'un coude de la Chicago River, coincé entre Michigan Avenue et Wabash Avenue. Et

donc ces quatorze personnes vivraient sous l'eau, ce qui a bien sûr attiré beaucoup de commentaires ironiques de la part d'humoristes et de journalistes.

En fait il ne s'agit pas d'une erreur de collecte mais de l'application de techniques dites de protection différentielle des données (*differential privacy*) afin que les données fournies par les personnes recensées ne puissent pas permettre leur identification, ce que les techniques de « pseudonymisation » ne peuvent pas garantir à 100 %. On ajoute un « bruit » aléatoire qui ne modifie pas les résultats agrégés mais rend pratiquement impossible l'identification des personnes recensées. C'est ainsi que quatorze personnes recensées se sont retrouvées affectées au bloc 1002 !

De nombreux utilisateurs des données du recensement disent non seulement que ces techniques produisent des résultats absurdes comme ceux du bloc 1002, mais qu'elles pourraient également limiter la publication, pour des raisons de confidentialité, des informations de base sur lesquelles elles s'appuient. De nombreux débats ont eu lieu, parfois au sein même de l'US Census Bureau, chargé des opérations de collecte et de diffusion des données des recensements. Certains se demandent si le système a été suffisamment testé : la plupart des changements majeurs apportés au recensement sont testés jusqu'avec une décennie d'avance. Or la confidentialité différentielle a été mise en place avec seulement quelques mois d'études et de tests. La publication des résultats en a encore été retardée à un moment où la publication de résultats était déjà passablement entravée par la pandémie de la Covid-19.

(11) La notion de « bloc de recensement » ne recoupe pas nécessairement celle de « city block » (pâté de maisons), en particulier dans les zones rurales.

## **Nous avons lu**

### **Hommages rendus à deux éminents statisticiens qui nous ont quitté récemment, Jean-Claude Deville et André Vanoli**

Jean-Claude Deville, ancien inspecteur général de l'Insee, décédé le 17 novembre 2021, était mondialement connu pour ses recherches en sondages qui ont permis de grandes avancées en la matière et ses travaux sur le calage, l'équilibrage ou le partage des poids ont eu un impact considérable sur le plan international. Le dernier numéro (3 mai 2022) de la revue de la SFdS, Statistique et Société, lui a rendu hommage à travers un [article de Pascal Ardilly](#).

André Vanoli, ancien directeur à l'Insee, décédé le 20 février 2022, avait été l'un des pionniers de la comptabilité nationale française et a été un des experts mondialement reconnus dans ce domaine, notamment en ayant été l'un des cinq membres du groupe inter-secrétariats (Nations Unies, Banque Mondiale, FMI, OCDE, Commission européenne) à l'origine de la version 1993 du Système de Comptes nationaux. Un hommage préparé par Anne Harrison, François Lequiller et Jean-Paul Milot sera prochainement publié dans la Review of Income and Wealth, revue de IARIW (*International Association for Research on Income and Wealth*).

On pourra aussi trouver des hommages à [Jean-Claude Deville](#) et à [André Vanoli](#) dans les pages « In Memoriam » du site de l'ISI (International Statistical Institute).

### **Un dossier dans le supplément Sciences et Médecine du quotidien Le Monde daté du 18 mai 2022 sur l'objet de recherche que constituent les modes de scrutin**

Cette question n'est pas neuve et on la fait remonter usuellement à Nicolas de Condorcet qui a probablement été, à la fin du XVIIIe siècle, le premier mathématicien à décrire scientifiquement les problèmes liés au système de vote uninominal (chaque électeur ne choisit qu'un seul nom dans l'ensemble des candidats). C'est le fameux « paradoxe de Condorcet » : un candidat préféré à n'importe quel autre candidat dans un « duel » peut très bien se trouver éliminé lors du scrutin uninominal ; il semblerait bien que, selon plusieurs enquêtes par sondage effectuées à l'époque, François Bayrou aurait été le « vainqueur de Condorcet » lors de l'élection présidentielle de 2007. Le dossier du journal *Le Monde* présente les travaux d'un groupe pluridisciplinaire de chercheurs issus de quatre universités françaises regroupés sous le nom collectif de *Voter Autrement* ; le 10 avril dernier, à l'occasion du premier tour de l'élection présidentielle, ce collectif a ainsi testé plusieurs modes de scrutin auprès de quelques centaines d'électeurs strasbourgeois invités à se présenter dans un second « bureau de vote » à la sortie du bureau officiel et à y exprimer leurs préférences selon d'autres modes de scrutin bien décrits dans le dossier du *Monde*. On constate que les divers modes de scrutin proposés peuvent donner des résultats très différents et que le mode de scrutin uninominal à deux tours largement utilisé en France souffre de nombreux défauts au niveau mathématique ainsi qu'au niveau sociétal.

On peut trouver surprenant - pour ne pas dire plus - que ce dossier n'ait pas aussi fait référence à l'un

des ouvrages parus dans la [collection « Le Monde des données » de la SFdS](#) éditée par EDP Sciences sous le titre « *Comment être élu à tous les coups ? Petit guide mathématique des modes de scrutin* ». Cet ouvrage présente une vue beaucoup plus complète que le dossier du Monde (il inclut notamment les scrutins de liste) ... et les auteurs ne manquent pas, eux, de mentionner les travaux des collectifs *Voter Autrement* et *Mieux voter* sur lesquels s'appuie le dossier du Monde. *La Statistique dans la Cité* en recommande vivement la lecture.

**Le Monde, dans son édition datée du vendredi 3 juin 2022, fait état du Rapport sur les riches en France publié par l'Observatoire des inégalités**

Dans ce rapport, cet observatoire, qui, soulignons-le, n'est pas malgré son nom, un organisme officiel, mais une association, définit un seuil de richesse. Ce seuil est fixé au double du niveau de vie médian, soit 3 673 euros par mois pour une personne seule, après impôts (5 511 euros pour un couple sans enfants, 7 700 euros pour une famille avec deux enfants). Selon cette définition la France compterait donc quatre millions et cinq cent mille « riches », soit 7,1 % de la population. Le seuil de richesse en patrimoine est fixé au triple du patrimoine médian : quatre millions et cinq cent mille de ménages (16 %) dépassent ce seuil évalué à 490 000 euros.

Ce rapport dresse aussi un portrait sociodémographique des « riches ». En moyenne, ils sont un peu plus âgés que l'ensemble de la population. Ce sont le plus souvent des cadres supérieurs du privé ou du public, souvent en couple et sans enfant à charge.

Le principal problème que pose ce rapport est qu'il semble proposer une symétrie entre *seuil de pauvreté* et *seuil de richesse*, notamment en raison du mode de calcul de ce dernier seuil. Or cette symétrie n'est qu'apparente ; en effet, la distribution des revenus n'est pas normale (au sens probabiliste de l'adjectif « normale »), mais est très largement asymétrique : il s'agit bien d'une distribution « en cloche », mais très largement aplatie vers sa droite, ce qui interdit un parallèle entre ces deux seuils. Et, alors qu'il existe une définition « officielle » du seuil de pauvreté (Eurostat par exemple), ce n'est pas le cas pour le « seuil de richesse » dont la définition est donnée par cet Observatoire. De même, il n'y a pas vraiment de symétrie entre le mot « riche » utilisé dans ce rapport et le mot « pauvre » associé aux seuils de pauvreté proposés par la statistique publique.

### ***Vie du groupe***

Au cours de l'assemblée générale de la SFdS tenue le mardi 14 juin à l'occasion des Journées de Statistique (JdS) à Lyon, les résultats des élections des membres du Conseil et des bureaux des groupes ont été annoncés. Pour ce qui concerne le groupe Statistiques et Enjeux Publics, les mandats de Chantal Cases, Marion Selz et Olivier Vasseur ont été reconduits pour une durée de trois ans.

Nous avons dû remettre à la saison 2022 - 2023 la séance prévue le mardi 12 avril sur la mesure de l'économie de la défense et des ventes d'armement pour lequel nous avons invité Christian Schmidt, professeur émérite à l'Université de Paris IX-Dauphine. Les séances des Cafés lyonnais de la statistique ont repris en « présentiel » et trois cafés ont pu ainsi être organisés au Café de la Cloche sur les thèmes suivants : sociologie de la quantification avec Fabrice Bardet de l'ENTPE, vote par évaluation avec Jean-François Laslier du CNRS et indice des prix à la consommation par Jean-Philippe Grouthier de l'Insee.

Il faut enfin souligner une grande première : la co-organisation le jeudi 16 juin par les équipes parisienne et lyonnaise des cafés, dans le cadre des événements satellites des 53e Journées de la Statistique tenues à Lyon, d'un *Café de la statistique* en mode « hybride » avec Cédric Philibert de l'Institut Français des Relations Internationales (Ifri) sur « *la transition énergétique : quels choix en France et en Europe ?* »

Responsable de l'infolettre : Chantal Cases, présidente du groupe SEP  
Rédacteur en chef : Jean-Louis Bodin  
Secrétaire de rédaction : Jean-Pierre Le Gléau  
Webmestre : Érik Zolotoukhine