

La statistique dans la cité n° 41 - octobre 2024

Sommaire du n°41 – octobre 2024

Éditorial

Méthodes :

- *Les sondages et l'élection présidentielle américaine du 5 novembre 2024*
- *Un problème d'échantillonnage*

Vie des institutions :

- *Que se passe-t-il en Inde ?*

Outils :

- *Filosofi : un nouvel ensemble d'indicateurs sur les revenus*

Humour :

- *Les dictateurs ne sont pas bons en maths*

Nous avons lu

Annonce

Vie du groupe

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

Éditorial

Après un été marqué par l'immense succès des Jeux olympiques et paralympiques, mais aussi par la poursuite et l'aggravation des conflits armés en Ukraine et au Proche-Orient, *La Statistique dans la Cité* reprend son rythme bimestriel.

Les quatre mois écoulés depuis la parution de notre précédent numéro n'auront pas été une trêve pour les instituts de sondage, de nombreuses élections ayant eu lieu ou étant prévues prochainement. Dans beaucoup de cas, les intentions de vote mesurées par ces instituts n'ont pas été très éloignées des résultats définitifs. Toutefois, l'élection présidentielle américaine est une de celles les plus difficiles à analyser pour ces instituts en raison du caractère très particulier de ce scrutin. Dans ce numéro, nous rappelons à ce sujet une brève publiée en décembre 2016 à propos du scrutin du mardi 8 novembre 2016 qui avait vu la victoire de Donald Trump alors qu'Hillary Clinton avait la majorité des voix au niveau fédéral. On trouvera aussi, sous la forme d'un billet d'humour, une démonstration de la fraude manifeste qui a permis la réélection de Nicolas Maduro à la présidence vénézuélienne en juillet dernier en faisant appel à un simple calcul probabiliste.

Nos lecteurs vont retrouver dans ce numéro les rubriques habituelles de *La Statistique dans la Cité*. Et comme à l'accoutumée, nous souhaitons qu'ils 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.

Méthodes

Les sondages et l'élection présidentielle américaine du 5 novembre 2024

À cause des particularités des modalités de l'élection présidentielle américaine (élection de grands électeurs dans chacun des cinquante États, avec la règle « *The winner takes all* »), le travail des instituts de sondage est particulièrement difficile et est de ce fait souvent critiqué. Les élections de 2016 et de 2020 n'avaient pas échappé à ces critiques et *La statistique dans la cité* avait publié à ce sujet deux brèves dans son numéro 22 de décembre 2020 sous le titre « *La fiabilité des sondages* ».

pré-électorales aux USA est-elle avérée ? » ainsi que dans son numéro 2 de décembre 2016 sous le titre « Les instituts de sondage américains ont-ils vraiment failli ? Creusons un peu ». À quelques jours du scrutin de 2024, il n'est pas inutile de rappeler le texte de cette brève de 2016 :

De fait, au niveau du nombre total de voix obtenues au niveau fédéral, les sondages sont loin d'avoir failli puisqu'il semble bien, d'après les derniers comptages, qu'Hillary Clinton ait obtenu plus de deux millions de voix de plus que son concurrent élu.

Mais bien entendu, compte tenu du système électoral américain, tout se joue au niveau des États, et surtout au niveau des quelque sept ou huit États clés (les fameux « Swing States »). Des sondages sont aussi bien sûr effectués dans les États et en particulier dans les « Swing States ». Or dans chacun de ces derniers, les sondages effectués les jours précédant le 8 novembre donnaient une chance sur deux de gagner à chacun des deux principaux candidats. Sans doute le raisonnement fait par tous les commentateurs a-t-il été le suivant : la probabilité que l'un ou l'autre des deux principaux candidats gagne chacun de ces États était de 50 % ; mais la probabilité qu'un des deux candidats gagne simultanément dans les huit États considérés était très faible (0,5 à la puissance 8) et donc, puisqu'il suffisait qu'Hillary Clinton gagne dans seulement deux ou trois de ces « Swing States » pour obtenir la majorité des grands électeurs au niveau fédéral, les commentateurs ont pu estimer que sa probabilité de gagner l'élection au niveau fédéral était sensiblement supérieure à 0,5. Mais ce raisonnement aurait supposé pour être exact que les événements dans les huit urnes que constituent les huit « Swing States », aient été indépendants. Ce qui n'était évidemment pas le cas, et la probabilité que l'un ou l'autre des deux candidats gagne simultanément tous ces États était donc en fait plus voisine de 0,5 que de zéro. Et donc les commentateurs auraient dû être prudents. Ce sont eux, plutôt que les instituts de sondage américains, qui ont failli.

De façon générale, des résultats différents de ceux que pourraient laisser apparaître les sondages peuvent se produire, avec tout système électoral, quand les intentions de vote sont proches les unes des autres, et ce ne sont pas les sondages qu'il faut incriminer mais l'utilisation qui en est faite, en particulier dans le cas américain.

Un problème d'échantillonnage

La région Ile de France a dépensé 250 millions d'euros lors des Jeux olympiques et paralympiques pour s'assurer que les transports parisiens seraient à la hauteur de l'événement. Cela a certainement réduit les temps d'attente des usagers et augmenté leur satisfaction. Le calcul de ce temps d'attente pose cependant un problème d'échantillonnage statistique intéressant.

Pour comprendre le problème, voici un petit exercice à l'attention de nos lecteurs. Supposons que les métros respectent un intervalle de 5' en moyenne entre deux départs du terminus. L'intervalle moyen entre deux métros sera donc de 5'. Quel est alors le temps moyen d'attente de l'utilisateur qui arrive dans une station ? La réponse se trouve dans le paragraphe suivant : réfléchissez un peu à la question avant de le lire.

Un raisonnement courant est le suivant : il n'y a pas de raison spécifique que le transporteur m'en veuille plus qu'à d'autres. Je vais donc de temps en temps arriver juste après le départ du précédent métro, de temps en temps arriver juste avant l'arrivée du métro suivant, et en général de manière uniforme entre les deux. Le temps d'attente sera donc la moitié de l'intervalle moyen entre deux métros, soit 2'30". Raisonnement tentant, mais faux.

On peut montrer que le temps d'attente est en fait de 5' (moyennant certaines hypothèses sur la distribution de la loi statistique régissant la durée entre deux métros). Tout simplement parce que la probabilité d'arriver lors d'une durée longue est plus importante que la probabilité d'arriver lors d'une durée courte, ce que négligeait le raisonnement ci-dessus.

En termes d'échantillonnage, si on tirait un échantillon des durées qui recouvrent votre arrivée dans la station, cet échantillon serait composé de durées en moyenne plus longues que la durée moyenne entre deux métros.

Ce phénomène d'échantillonnage endogène (ou stock sampling) est connu depuis bien longtemps. Un phénomène bien documenté est qu'un échantillon de chômeurs extrait à l'instant t de registres de Pôle Emploi est biaisé : les durées plus longues sont surreprésentées.

Vie des institutions

Que se passe-t-il en Inde ?

Le 10 septembre dernier, le ministre indien des statistiques et de la mise en œuvre des programmes (MoPSI) a brusquement dissous le Comité permanent des statistiques (SCoS) qui avait été créé en décembre 2019 pour superviser toutes les enquêtes statistiques du ressort de ce ministère. Sa

présidence avait été confiée au Dr Pronab Sen, qui avait été le premier à recevoir le titre de Chef statisticien de l'Inde, fonction qu'il avait exercée de 2007 à 2010. La raison officielle de cette décision a été que ce Comité faisait double emploi avec un Comité permanent des enquêtes nationales par sondage (SCNSS) créé seulement au mois de juin dernier. Que se cache-t-il derrière cette dissolution ? Plusieurs médias indiens ont noté qu'elle est intervenue alors que plusieurs membres du SCoS avaient fait part de leurs inquiétudes concernant le retard important dans la mise en route du recensement décennal qui était prévu en 2021, ce qui pouvait avoir des conséquences sur la fiabilité des enquêtes par sondage. Un article sur le bilan du président Narendra Modi paru dans *Le Monde* daté du 19 septembre 2024 faisait écho à ces inquiétudes. La création d'un nouveau comité en juin dernier n'était-elle pas en fait la réponse du gouvernement indien à ces inquiétudes et n'ouvrait-elle pas la voie à la décision de dissoudre le SCoS ? Elle interroge en tout cas sur la crédibilité qu'il faut accorder aujourd'hui au système statistique indien qui jouissait autrefois d'une excellente réputation en Asie et dans le monde.

Outils

Filosofi : un nouvel ensemble d'indicateurs sur les revenus

Le Centre d'accès sécurisé aux données (CASD) vient de mettre à disposition sous l'acronyme de Filosofi un nouveau dispositif préparé par l'Insee et intitulé [Fichier Localisé Social et Fiscal](#) qui permet d'accéder à un ensemble d'indicateurs sur les revenus déclarés et sur les revenus disponibles après redistribution et imputation de revenus financiers non déclarés, à l'échelle communale, supra-communale et infra-communale. Il s'agit des indicateurs usuels d'analyse de la distribution des revenus par unité de consommation, mais aussi des taux de pauvreté monétaire et d'indicateurs de structure des revenus.

Humour

Les dictateurs ne sont pas bons en maths

C'est ce qu'annonce Étienne Ghys, secrétaire perpétuel de l'Académie des sciences, dans sa chronique Carte Blanche du supplément *Sciences et Médecine* du *Monde* daté du 11 septembre dernier. Un de ses anciens doctorants lui a envoyé un argument mathématique qui confirme l'élection frauduleuse de Nicolas Maduro à la présidence du Venezuela le 25 juillet dernier. Le Conseil national électoral a annoncé qu'il a recueilli 5 150 092 voix, soit exactement 51,2 % des 10 058 774 suffrages exprimés, tandis qu'Edmundo Gonzalez, le candidat de l'opposition, en aurait recueilli 4 445 978, soit exactement 44,2 %. Ces pourcentages sont trop parfaits ! En effet, en les appliquant au nombre des suffrages exprimés, on retombe exactement sur le nombre d'électeurs de chacun des candidats, ce qui laisse penser que ces pourcentages ont été choisis au hasard en vue de produire un résultat « acceptable », et que le Conseil électoral aurait ensuite calculé le nombre de votes sans se préoccuper du fait que la probabilité que ces pourcentages soient tombés juste est pratiquement nulle (une fois sur cent millions dit Étienne Ghys, cinq fois moins que de gagner au Loto). Un fraudeur intelligent aurait modifié le nombre des votes de chaque candidat de quelques milliers d'unités pour faire « plus vrai ».

Étienne Ghys rappelle que, lors d'une élection en Russie il y a une dizaine d'années, des manifestants avaient brandi des pancartes « *Nous croyons Gauss* », en s'appuyant sur la forme bizarre du graphique du nombre de bureaux de votes en fonction du pourcentage de participation, très éloigné de la fameuse courbe en cloche.

Décidément, les dictateurs ne sont pas bons en calcul des probabilités.

Nous avons lu

La division de la population des Nations Unies a publié de nouvelles projections de population mondiale le 11 juillet dernier

Ces projections sont analysées par l'Ined dans un [numéro de son magazine Focus](#) paru en juillet dernier. D'après cette analyse, la planète compte aujourd'hui 8,2 milliards d'habitants et devrait en compter 9,7 en 2050 puis 10,3 dans les années 2080 pour commencer à diminuer et atteindre 10,2 milliards à la fin du siècle. Toutefois la population de l'Afrique subsaharienne augmenterait des deux-tiers d'ici 2050 alors que celle de l'Europe et de l'Amérique du Nord resterait au même niveau qu'aujourd'hui. L'indice synthétique de fécondité mondial qui est passé de 3,31 enfants par femme en

1990 à 2,25 aujourd'hui devrait encore reculer à 2,07 en 2050 mais serait encore de 4,3 en Afrique subsaharienne.

L'espérance de vie à la naissance est passé de 64,0 ans en 1990 à 73,3 ans aujourd'hui ; elle devrait atteindre 77,0 ans en 2050, mais les écarts entre pays, bien qu'ils se soient considérablement réduits restent importants et les pays les moins avancés connaissent un retard de 7 ans par rapport à la moyenne mondiale.

Avec un indice synthétique de fécondité qui n'est plus que d'un seul enfant par femme, la Chine va connaître un effondrement de son chiffre de population (640 millions en 2100 contre 1 421 millions aujourd'hui) tandis que celle de l'Inde, qui a dépassé cette année celle de la Chine (1 444 millions), continuera à croître pour atteindre 1,5 milliards à la fin du siècle.

Les migrations sont devenues une composante majeure de la croissance démographique dans beaucoup de pays développés.

Évaluation des impacts des Jeux de Paris

Dans son édition du 9 septembre dernier, le Webmagazine Variances, édité par l'association des anciens élèves de l'Ensaé, propose une [analyse des résultats financiers des Jeux Olympiques et Paralympiques](#) organisés cet été à Paris qui conclut qu'ils ont été peu chers comparés aux éditions précédentes et « gratuits » pour les finances publiques. Leur coût total est estimé, d'après l'auteur, à 11,8 milliards d'Euros, répartis à peu près par tiers entre le Comité d'organisation des Jeux Olympiques (Cojop), la Société de livraison des ouvrages olympiques (Solideo) et des dépenses diverses, notamment pour la sécurité. Si on compare les dépenses totales à celles des autres Jeux depuis 1980 (en Euros 2024 et en corrigeant en fonction de l'évolution du nombre d'épreuves qui n'a cessé d'augmenter), seuls les Jeux d'Atlanta en 1988 et ceux de Sydney en 1992 ont coûté moins cher, les Jeux les plus coûteux ayant été ceux de Pékin en 2008 et ceux de Rio en 2016 (plus de 40 milliards d'Euros). Sur cette dépense totale de 11,8 milliards, il y aurait 5,2 milliards d'argent public (en particulier les dépenses du Cojop auront été couvertes à 96 % par la billetterie, les parrainages et la subvention du CIO). Mais les rentrées fiscales (TVA, cotisations salariales et patronales sur les 137 000 emplois créés, ...) sont estimées par l'auteur à 5,3 milliards d'Euros. Ainsi, les recettes publiques générées par les dépenses engagées pour l'organisation des Jeux seraient globalement équivalentes aux dépenses publiques réalisées.

Dans un autre registre, l'Insee, dans son billet de blog [« Impacts économiques et sociaux des Jeux olympiques et paralympiques : à vos stats, prêts, partez »](#), met l'accent sur la mobilisation de l'institut et du service statistique ministériel en charge de la jeunesse et des sports pour évaluer les impacts sportifs, économiques, sociaux et environnementaux des Jeux. De nombreuses enquêtes ont été lancées, pour juger si les objectifs de ces Jeux ont été atteints (au-delà du nombre de médailles !). Ont-ils contribué au dynamisme économique et à l'attractivité touristique de l'Île-de-France et de l'ensemble du pays ? Permettront-ils d'encourager la pratique sportive ? Quel sera leur impact environnemental ? L'évaluation économique a déjà commencé et se poursuivra sur plusieurs années : emploi mobilisé pour la livraison des ouvrages olympiques, étude des flux touristiques, impact sur le PIB, transformation socio-économique de la Seine-Saint-Denis.

Les Jeux ont aussi fourni l'occasion d'enrichir les connaissances et les dispositifs de suivi de la pratique sportive. L'Injep (Institut national de la jeunesse et de l'éducation populaire) a amélioré ses enquêtes statistiques, en adoptant des dispositifs de suivi plus fréquents, comme le [Baromètre national des pratiques sportives](#) désormais annuel ou la mise en place d'un indicateur avancé de l'évolution des licences sportives. Il a utilisé des enquêtes de plusieurs services statistiques ministériels pour mieux appréhender les pratiques sportives des personnes handicapées. Le service statistique du ministère de l'Enseignement supérieur et de la recherche a entrepris des travaux pour mieux connaître les étudiants sportifs de haut niveau et le recensement des équipements sportifs, suspendu depuis 2021, a été relancé par le ministère en charge du sport. La charte en faveur de l'Emploi et du développement territorial a conduit la Solideo (Société de livraison des ouvrages olympiques) à créer un dispositif inédit de suivi des personnes en insertion sur les chantiers des Jeux.

Annonce

L'Association Internationale pour la Statistique Officielle (IAOS), l'une des associations de l'Institut International de Statistique (ISI), annonce l'ouverture des candidatures au Prix du Jeune Statisticien Officiel pour l'année 2025. Les candidats et candidates doivent travailler dans une organisation de la statistique publique et avoir moins de 35 ans. Le ou la titulaire du Premier Prix recevra une bourse de 1 500 €, une invitation à présenter son travail à une conférence statistique internationale de son choix

(avec prise en charge de ses frais de transport et de séjour), la publication de son papier dans le *Statistical Journal of the IAOS* et deux années de cotisation à l'IAOS. Seront aussi décernés un deuxième et un troisième prix, ainsi qu'un prix spécial attribué à un candidat d'un pays en développement. Les modalités précises de candidature sont affichées sur le site de l'IAOS à l'adresse <https://iaos-isi.org/ysp/>. Une affiche est également [disponible en français](#).

En 2019, le prix avait été attribué à Vianney Costemalle, administrateur de l'Insee, pour son papier [“Detecting geographical differencing problems in the context of spatial data dissemination”](#).

La Statistique dans la Cité invite ses lecteurs à identifier de possibles candidats ou candidates à ce Prix et à soumettre leur proposition avant le 21 février 2025.

Vie du groupe

Depuis la parution du précédent numéro de *La statistique dans la cité*, un Café de la statistique a été organisé au Café du Pont-Neuf le mardi 8 octobre où Sylvie Lambert de l'École d'économie de Paris a traité de la mesure de la pauvreté et de son rôle pour les politiques de développement.

Les Cafés suivants auront lieu :

- le mardi 19 novembre : l'efficacité de la réponse pénale avec Pascal Chevalier, chef du SSM du ministère de la Justice ;
- le mardi 10 décembre : les classes moyennes avec Thomas Amossé, membre du Laboratoire interdisciplinaire pour la sociologie économique (Lise) du CNAM ;
- le mardi 14 janvier 2025 : le RESIL (Répertoire statistique des individus et des logements) avec Olivier Lefebvre (Insee) ;
- le mardi 11 février 2025 : la sécurité alimentaire avec Akiko Suwa-Eisenmann, chercheuse senior à l'INRAE (Institut national de la recherche pour l'agriculture, l'alimentation et l'environnement).

Notre groupe aura aussi été associé à l'hommage rendu à notre ancienne présidente, Chantal Cases, décédée il y a quinze mois, au Centre des colloques du Campus Condorcet le 18 octobre.

Les lecteurs de *La statistique dans la cité* sont invités à proposer des thèmes de Café en écrivant à l'adresse du Groupe sep@sfds.asso.fr.

Responsable de l'infolettre : Antoine Moreau, président du groupe SEP

Rédacteur en chef : Jean-Louis Bodin

Secrétaire de rédaction : Jean-Pierre Le Gléau

Webmestre : Érik Zolotoukhine