



**LES ENQUETES DE PARTICIPATION ELECTORALE : UNE COMBINAISON
ORIGINALE DE DONNEES ADMINISTRATIVES ET D'ENQUETE**

Elisabeth Algava (), Kilian Bloch(*)*

() Insee, DSDS, Division Enquêtes et études démographiques*

elisabeth.algava@insee.fr, kilian.bloch@insee.fr

Mots-clés : Données administratives, Collecte, Élections, Participation électorale, Abstention

Domaine concerné : Collecte, Multimode

Résumé

La participation aux élections est un indicateur de l'implication des citoyens vis-à-vis des institutions et des grands débats collectifs. Depuis 1983, l'Insee réalise des enquêtes de participation électorale (EPE) lors de chaque élection présidentielle ou législative. Ces enquêtes permettent l'étude approfondie des déterminants sociodémographiques des comportements de vote et de leur évolution d'une élection à l'autre. Diffusées via l'ADISP (Archives de Données Issues de la Statistique Publique) et plus récemment également via le CASD (Centre d'Accès Sécurisé à Distance), elles ont donné lieu à des publications de la statistique publique mais aussi à des articles de recherche.

Le principe de ces enquêtes est de relier l'information sur la participation aux scrutins, à partir des listes d'émargements, et les caractéristiques individuelles figurant dans l'échantillon démographique permanent (EDP). Chaque personne enquêtée appartient à l'EDP. Celui-ci est un panel d'individus constitué par l'Insee depuis 1968. Il consiste à collecter et conserver des informations puisées dans plusieurs sources, qui sont des enquêtes ou des données administratives. Ce panel ne concerne que des individus nés certains jours de l'année. Les personnes sélectionnées pour l'EPE doivent en outre avoir été recensées récemment (en 2015 pour l'enquête de 2017, en 2020 pour l'enquête de 2022). Cela garantit la fraîcheur des informations sur les caractéristiques socio-démographiques des individus, par exemple en termes de situation familiale, de diplôme ou d'emploi. La consultation du répertoire électoral unique (REU) permet de vérifier que les personnes sélectionnées sont inscrites sur les listes électorales, et de connaître le lieu de leur inscription.

L'information sur la participation électorale est ensuite collectée directement dans les listes d'émargements, issues des bureaux de vote. Ces listes sont mises à disposition de tous les électeurs dans les locaux des préfectures dans les 10 jours qui suivent le scrutin électoral. Ce droit d'accès est prévu par le Code électoral (art. L68). La collecte est assurée par des agents des établissements régionaux de l'Insee qui relèvent la participation des individus à chacun des scrutins de l'élection présidentielle puis des élections législatives.

Pour l'édition de 2022, la collecte sera facilitée grâce au remplacement du Fichier général des électeurs (FE) par le REU, initié en 2016. D'abord, les listes électorales sont désormais directement issues du REU, ce qui évite les incohérences (rares) entre l'échantillon des personnes pour lesquelles la participation électorale est à rechercher et la liste des personnes figurant sur la liste d'émargement. Ensuite, le bureau de vote est enregistré dans le REU. Pour les éditions précédentes, des recherches préalables auprès des grandes communes étaient nécessaires pour connaître les bureaux de vote des électeurs, afin de retrouver plus facilement ces derniers dans les listes d'émargement. Enfin, les procurations sont aussi enregistrées dans le REU, ainsi il sera inutile de les recueillir « manuellement » sur les listes comme cela a été le cas en 2017.

Le plan de sondage est adapté aux enjeux de l'enquête : limiter le temps de la collecte, avoir une représentativité au niveau régional, enquêter aussi les personnes inscrites dans une commune différente de leur commune de résidence. Par rapport à l'enquête de 2017, la taille de l'échantillon sera plus réduite en 2022 (35 000 individus au lieu de 45 000), la représentativité étant recherchée à l'échelle des régions actuelles et non plus des anciennes régions.

Au total, l'enquête rend possible l'étude des comportements électoraux d'une population représentative des inscrits sur les listes électorales, avec une grande richesse de descripteurs socio-démographiques. Son protocole original, qui combine des informations déjà collectées à une information publique (listes d'émargement) sans jamais entrer en contact avec un enquêté pour le questionner, permet d'éviter les défauts inhérents aux enquêtes par questionnaire (non-réponse, biais de désirabilité, etc.). La faible quantité d'information à relever permet d'avoir un échantillon d'assez grande taille, de mener l'opération sur un temps limité et pour un coût modéré.

Abstract

Since 1983, INSEE has conducted voter participation surveys during each presidential or legislative election. They allow the study of the social determinants of voting and their evolution.

The principle of these surveys is to link information on voter participation, collected directly from the lists of voters gathered in the prefecture after each election, with the individual characteristics contained in the permanent demographic sample (EDP), a panel of individuals set up by INSEE since 1968.

The sampling plan is adapted to the challenges of the survey: to limit the collection time, to be representative at the regional level, and to survey people registered in a municipality different from their municipality of residence.

Article

Depuis plusieurs décennies, les enquêtes réalisées par l'Insee donnent la mesure des inégalités sociales dans la participation électorale. Cet article présente leurs apports et leur méthode, en insistant plus particulièrement sur l'édition de 2022, dont la collecte commencera quelques semaines après les JMS.

Ces enquêtes, réalisées par l'Insee depuis près de quarante ans, permettent de connaître le profil des électeurs qui participent à tous les scrutins systématiquement, de ceux qui s'abstiennent toujours et de ceux qui votent de façon intermittente, c'est-à-dire uniquement à certains tours de scrutin.

Elles reposent sur la sélection d'un échantillon de personnes inscrites sur les listes électorales et recensées récemment. Pour ces personnes, des agents des directions régionales de l'Insee relèvent sur les listes d'émargement la participation à chaque tour de scrutin.

1. Quarante ans d'enquêtes sur la participation électorale

L'Insee réalise des enquêtes de participation électorale depuis presque quarante ans.

1.1 Un principe commun : étudier la participation électorale de personnes de l'EDP, recensées récemment

L'idée de ces enquêtes est la même depuis leur origine : pour un échantillon de personnes faisant partie de l'échantillon démographique permanent (EDP), recensées récemment et inscrites sur les listes électorales, la consultation des listes d'émargement en préfecture permet de relever si la personne a participé ou non à chaque tour des scrutins étudiés (**figure 1**). Ainsi pour l'enquête expérimentale de 1983, « l'échantillon a été construit en prenant en compte les personnes figurant dans l'échantillon démographique permanent, donc recensées en 1975, et portées dans le fichier électoral en mars 1983 avec une commune d'inscription » **[1]**.

L'EDP est un panel sociodémographique de grande taille mis en place en France, pour étudier la fécondité, la mortalité, les parcours familiaux, les migrations géographiques au sein du territoire national, la mobilité sociale et la mobilité professionnelle, les carrières salariales et les niveaux de vie ainsi que les interactions possibles entre ces différents aspects **[2]**. Le principe général consiste à cumuler pour les individus appartenant à l'échantillon (environ 4 % de la population) des informations collectées dans différentes sources statistiques. Depuis sa création en 1968, l'EDP intègre :

- 1) les bulletins d'état civil de naissance, de mariage, de décès depuis 1968 ;
- 2) les recensements exhaustifs de 1968, 1975, 1982, 1990 et 1999 puis les enquêtes annuelles de recensement à partir de 2004 ;
- 3) des informations sur l'inscription électorale depuis les années 2000 à partir du fichier électoral, et à partir de 2020 du répertoire électoral unique (REU), mis en place en 2016.

D'autres sources sont venues enrichir l'EDP au fil des années **[3]** : le panel "tous salariés" depuis 1967, qui contient des informations sur les rémunérations perçues ; depuis 2011, les données sur les revenus, les aides perçues, les impôts et les pensions versées par le ménage de l'individu, grâce au fichier démographique sur les logements et les individus (Fidéli) et au fichier sur les revenus localisés sociaux et fiscaux (FiLoSoFi). Prochainement, des données sur les non-salariés vont être intégrées à l'EDP.

Du point de vue des enquêtes de participation électorale, l'essentiel est que grâce à l'EDP il est possible de sélectionner un échantillon d'individus inscrits sur les listes électorales, nés un jour EDP et recensés une année donnée.

En effet, la gestion de l'EDP intègre un identifiant technique permettant de faire le lien entre l'EDP et le RNIPP (répertoire national d'identification des personnes physiques). Ce lien est indispensable au suivi du panel EDP, mais il ne figure bien entendu pas dans les fichiers statistiques à destination des chargés d'études ou des chercheurs. Le REU (et le fichier électoral antérieurement) comprend également un tel lien, nécessaire lui aussi à des fins de gestion. L'échantillon de l'EPE est donc choisi dans le REU, parmi les personnes nées un jour EDP et recensées récemment.

L'échantillon ainsi constitué comprend uniquement les informations du REU nécessaires à la collecte : noms, prénoms, date de naissance et lieu d'inscription. Ces informations servent à rechercher les personnes sur les listes d'émargement et à relever leur participation aux différents tours de scrutin.

Une fois la collecte achevée, les noms et prénoms des personnes de l'échantillon sont détruits. Ils ne sont donc pas rapprochés des données statistiques de l'EDP. Ils ne servent qu'à la consultation des listes d'émargements.

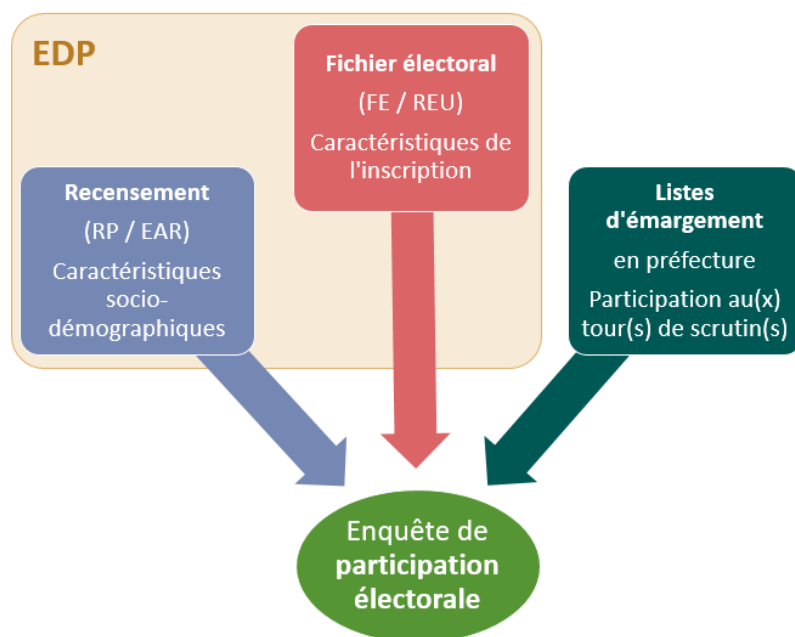
Seule l'information sur la participation au vote est conservée pour être ensuite rapprochée des informations statistiques de l'EDP portant sur la caractérisation socio-démographique des personnes. Ce rapprochement est fait dans le cadre du traitement de l'enquête participation électorale et afin de réaliser des études statistiques.

La consultation des listes d'émargement en préfecture permet de savoir si la personne a voté à chaque tour d'élection. Cette consultation est prévue dans le code électoral, article L68 :

« Tant au premier tour qu'éventuellement au second tour de scrutin, les listes d'émargement de chaque bureau de vote, ainsi que les documents qui y sont réglementairement annexés, sont joints aux procès-verbaux des opérations de vote transmis immédiatement après le dépouillement du scrutin à la préfecture. [...] Sans préjudice des dispositions de [l'article L0. 179](#) du présent code, **les listes d'émargement déposées à la préfecture ou à la sous-préfecture sont communiquées à tout électeur requérant pendant un délai de dix jours à compter de l'élection** et, éventuellement, durant le dépôt des listes entre les deux tours de scrutin, soit à la préfecture ou à la sous-préfecture, soit à la mairie. »

La seule condition est donc que les agents de l'Insee chargés de la consultation soient inscrits sur les listes électorales.

Figure 1 : Le principe des enquêtes de participation électorale



1.2 Les principales évolutions des enquêtes

Toutes les enquêtes reposent sur le même principe, mais elles ont connu quelques évolutions en quarante ans.

Première évolution, à partir de 1988, les enquêtes suivent pour le même échantillon d'électeurs, constitués en un panel, la participation à plusieurs scrutins. En 1988-1989, l'échantillon est suivi durant 7 tours au maximum : élections présidentielle, législatives, municipales et référendum sur la Nouvelle Calédonie (**figure 2**). Depuis l'édition de 2012, l'Insee réalise des enquêtes de participation électorale

tous les 5 ans, lors des élections présidentielle et législatives. Les électeurs sont donc suivis durant 3 ou 4 tours de scrutin.

Seconde évolution, la couverture géographique de l'enquête a été progressivement étendue : la première enquête expérimentale de 1983 portait sur les grandes villes de France métropolitaine, puis l'enquête a couvert toute la métropole. En 2012, trois départements d'Outre-mer ont été intégrés de façon expérimentale à l'enquête : la Martinique, la Guadeloupe et La Réunion, mais les résultats ont été publiés séparément. En 2017, la méthodologie de réalisation de l'enquête dans ces trois départements a été identique aux autres et des résultats conjoints ont été publiés. En outre, depuis 2017 l'enquête est aussi réalisée en Guyane, et le sera également à Mayotte pour la première fois en 2022. L'ensemble des données sera intégrée à la base statistique nationale.

Troisième évolution importante, la taille de l'échantillon est quintuplée (pour atteindre environ 40 000 inscrits) à partir de l'édition 1995-1997, afin de pouvoir diffuser des résultats régionaux. Lors de l'édition de 2017, il était prévu de maintenir une représentativité au niveau des 22 anciennes régions de métropole, la loi modifiant le nombre de régions métropolitaines étant alors très récente (loi de 2015 entrée en vigueur au 1^{er} janvier 2016). Mais l'enquête de 2017 a été analysée uniquement selon les contours des nouvelles régions, les seules en vigueur lors de la publication des études. Pour entériner la diminution du nombre de régions en métropole, l'échantillon de l'enquête de 2022 a été réduit à 35 000 inscrits.

En 2022, les résultats détaillés pourront être publiés pour 16 régions : les 13 régions de métropole, la Martinique, la Guadeloupe et La Réunion. En Guyane et à Mayotte, le nombre de personnes faisant partie de l'EDP et recensées récemment est trop faible pour constituer un échantillon exploitable isolément. Mais leur inclusion est importante pour l'élaboration de données nationales couvrant l'ensemble du territoire.

Figure 2 : L'évolution des enquêtes de participation électorale

Dates	Scrutins concernés	RP source d'info	Échantillon	Couverture	Représentativité
1983	Municipales	1982	3000	Grandes villes, France métro	Grandes villes
1984	Européennes	1982	7000	France métro	France métro
1988	Présidentielle et Législatives Référendum sur la Nouvelle Calédonie	1982	8000	France métro	France métro
1989	Municipales				
1995 1997	Présidentielle, Municipales Législatives	1990	39 000	France métro	Régions
2002 2004	Présidentielle et Législatives Régionales et Européennes	1999	39 000	France métro	Régions
2007 2008	Présidentielle et Législatives Municipales	1999	40 000	France métro	Régions
2012	Présidentielle et Législatives	Métro : 2010 DROM : 2006-2010	45 000	France entière + 3 DROM*	Régions
2017	Présidentielle et Législatives	2015	45 000	France entière + 4 DROM	Régions (sauf Guyane)
2022	Présidentielle et Législatives	2020 2021 pour Mayotte	35 000	France métro + 5 DROM	Régions 2015 (sauf Guyane et Mayotte)

* enquête expérimentale.

1.3 Les objectifs de l'EPE 2022 : pourquoi des enquêtes de participation électorale à l'Insee ?

1.3.1 Une grande richesse d'information

Le ministère de l'Intérieur publie à l'issue de chaque tour de scrutin des résultats détaillés : le nombre de voix obtenues par chaque candidat (ou, selon les scrutins, par liste ou par choix), le nombre d'inscrits, d'abstentions, de votants, de votes blancs, nuls et exprimés (**figure 3**). Ces résultats sont publiés par région et département et les résultats par bureaux de vote sont diffusés sur le site data.gouv.fr¹. Les enquêtes de participation électorale ne permettent pas de savoir pour qui les électeurs ont voté. Qu'apportent alors de plus les enquêtes de l'Insee ?

Figure 3: les résultats du second tour de l'élection présidentielle publiés par le ministère de l'Intérieur

Résultats de l'élection présidentielle 2017

23 avril et 7 mai 2017



[Accueil Présidentielle 2017](#)

France Entière

* Résultats au 2^d tour

Liste des candidats	Voix	% Inscrits	% Exprimés
M. Emmanuel MACRON	20 743 128	43,61	66,10
Mme Marine LE PEN	10 638 475	22,36	33,90

	Nombre	% Inscrits	% Votants
Inscrits	47 568 693		
Abstentions	12 101 366	25,44	
Votants	35 467 327	74,56	
Blancs	3 021 499	6,35	8,52
Nuls	1 064 225	2,24	3,00
Exprimés	31 381 603	65,97	88,48

En raison des arrondis à la deuxième décimale, la somme des pourcentages peut ne pas être égale à 100%.

Source : [https://www.interieur.gouv.fr/Elections/Les-resultats/Presidentielles/elecresult__presidentielle-2017/\(path\)/presidentielle-2017/FE.html](https://www.interieur.gouv.fr/Elections/Les-resultats/Presidentielles/elecresult__presidentielle-2017/(path)/presidentielle-2017/FE.html)

Les enquêtes de participation électorale offrent deux atouts importants pour l'étude de la participation et de l'abstention :

- le suivi en panel sur plusieurs scrutins

¹ Par exemple pour le premier tour de l'élection présidentielle de 2017 : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-des-23-avril-et-7-mai-2017-resultats-definitifs-du-1er-tour-par-bureaux-de-vote/>

Depuis 2012, les EPE portent uniquement sur la présidentielle et les législatives, qui ont lieu la même année depuis 2002. Ce sont ainsi 3 ou 4 tours qui seront analysés et permettent de caractériser le profil de chaque électeur : vote systématique s'il a voté à chaque tour, abstention systématique dans le cas contraire, et vote intermittent s'il n'a participé qu'à certains tours des scrutins.

- la caractérisation sociodémographique de l'électeur

Grâce aux informations issues de l'EDP, comme ils ont été recensés récemment, les électeurs peuvent être caractérisés de façon détaillée, bien au-delà de leur âge et leur sexe, par leur leur niveau de diplôme, leur profession, des informations sur leur lieu de résidence, leur situation familiale, les caractéristiques de leur conjoint et de leurs enfants, etc.

Les recensements apportent une grande richesse d'informations, mais leur principale limite est le délai entre la collecte des informations sociodémographiques et la collecte des informations sur la participation électorale. En deux ans, a fortiori en cinq ans, la situation conjugale et d'emploi des enquêtés a pu évoluer, surtout s'ils sont jeunes au moment de leur recensement, et au seuil des étapes d'entrée dans la vie adulte, ou encore s'ils sont proches de l'âge de la retraite. Depuis que le recensement n'est plus exhaustif mais collecté par enquête annuelle sur un échantillon d'adresses, l'information sociodémographique des EPE est toutefois plus récente puisque les personnes sont échantillonnées parmi celles ayant été recensées lors de l'enquête annuelle de recensement (EAR) de janvier-février N-2. En 2002, l'information provenait du recensement de 1999. En 2007 aussi : la profession, le diplôme, la situation familiale des personnes avaient pu changer en 8 ans. En 2022, l'information sera plus fraîche, puisqu'elle proviendra de l'EAR 2020. Comme en 2012 et en 2017, les caractéristiques socio-démographiques seront celles observées deux ans avant les élections.

En plus du recensement, des informations complémentaires issues de l'EDP ont été ajoutées à partir de l'EPE 2017 pour caractériser les électeurs, notamment le niveau de vie d'après les informations socio-fiscales récemment introduites dans l'EDP.

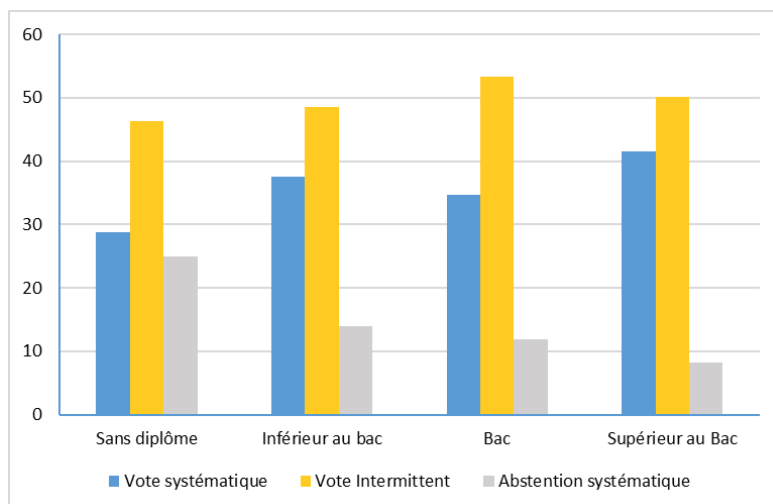
Ces deux atouts se combinent avec la possibilité de faire des exploitations régionales et de comparer les résultats d'une élection nationale à l'autre (**figures 4 et 5**).

Par rapport aux résultats détaillés publiés par le ministère de l'Intérieur, les enquêtes de l'Insee sur la participation électorale apportent donc un éclairage riche et complémentaire, et ce bien qu'elles ne portent que sur le fait de voter ou de ne pas voter, mais en aucun cas sur les candidats pour qui les électeurs ont voté. Elles soutiennent aussi la comparaison par rapport aux sondages électoraux, comme l'ont montré Braconnier, Coulmont et Dormagen [4] qui mettent en avant cinq avantages de ces enquêtes par rapport aux sondages réalisés par ailleurs : la représentativité de la population électorale, du fait que les individus appartenant à l'échantillon de l'enquête sont tirés aléatoirement, sans biais d'échantillonnage ; la grande taille des échantillons, permettant des analyses plus fines ; la richesse des informations disponibles (participation au vote, mais aussi descripteurs socio-démographiques des électeurs notamment) ; recueil de la participation sans biais grâce à la consultation des listes d'émargement ; approche longitudinales possibles, les enquêtes étant réalisées différentes années avec une méthodologie proche.

Les enquêtes de participation électorale constituent donc un dispositif d'enquêtes qui offre de nombreuses possibilités d'utilisation. Elles ont été de longue date utilisées pour mesurer les différences sociales de participation [5].

Lors de l'enquête de 2017, une opération coordonnée a été menée entre la division enquête et étude démographique (EED) responsable de l'opération, et les directions régionales, afin de publier des résultats simultanément au niveau national et au niveau régional. Deux Insee Première ont été publiés, le premier portant sur les résultats de 2017 [6] et le second sur l'évolution des différentiels sociaux de participation [7]. Quinze publications régionales ont été réalisées. L'ensemble des publications est diffusé sur [insee.fr \(https://www.insee.fr/fr/information/3142242\)](https://www.insee.fr/fr/information/3142242).

Figure 4 : En 2017, les moins diplômés votent moins (graphique des auteurs d'après [6])



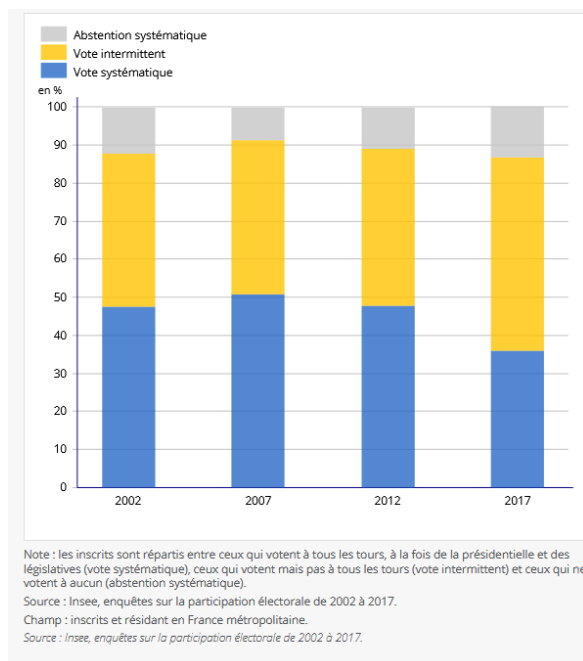
Note : les inscrits sont répartis entre ceux qui votent à tous les tours, à la fois de la présidentielle et des législatives (vote systématique), ceux qui votent mais pas à tous les tours (vote intermittent) et ceux qui ne votent à aucun (abstention systématique).

Source : Insee, enquête sur la Participation électorale 2017.

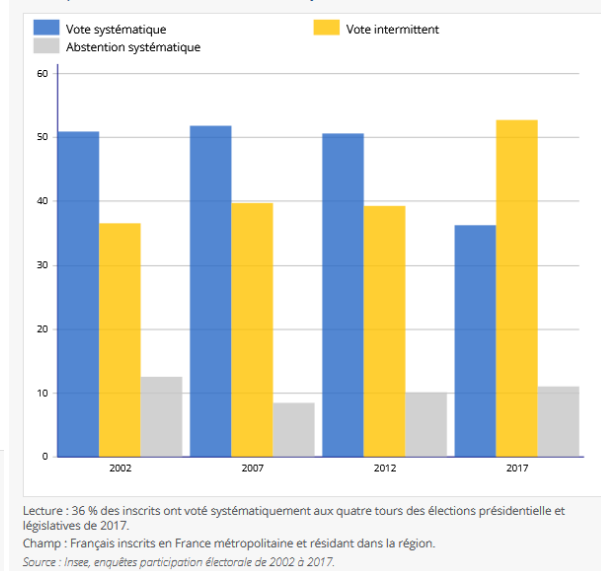
Champ : Inscrits et résidant en France hors Mayotte.

Figure 5 : Moins de vote systématique en 2017

Résultats nationaux [7]...



... et régionaux, comme en Pays de la Loire [8]



Des chercheurs en sciences politiques utilisent également régulièrement les enquêtes de participation électorale, pour analyser par exemple l'évolution des déterminants sociaux du vote. Braconnier, Coulmont et Dormagen ont ainsi montré que les taux élevés d'abstention observés en 2017 ne réduisaient pas les différentiels sociaux, bien au contraire [4]. Ainsi la participation des cadres au 1^{er} tour de la présidentielle était de 6 points supérieure à celle des ouvriers en 2007, 7 points en 2012 et 13 en 2017. Les auteurs concluent que « la progression de l'abstention au cours de la dernière décennie opère donc bien comme un amplificateur des inégalités de participation électorale. Le corps

des votants était moins représentatif de l'électorat dans son ensemble en 2017 qu'en 2012 et moins représentatif en 2012 qu'en 2007 ». Tiberj analyse pour sa part la participation en suivant des cohortes d'une élection à l'autre. Il en conclut que « Renouveau générationnel et intermittence du vote sont très liés. Les cohortes 1928-1937, 1938-1947 et 1948-1957 (celles du baby-boom) se caractérisent par une tendance à voter de manière constante, mais plus une cohorte est récente, moins elle compte d'électeurs constants et de plus en plus d'électeurs intermittents. En 2017, ils représentent plus de 50 % des cohortes 1958-1967, 1968-1977, 1978-1987, 1988 et après. » [9]

1.3.2 Des informations de plus en plus détaillées sur les modalités de la participation électorale

Grâce à la collaboration avec des chercheurs spécialistes du domaine, les EPE se sont enrichies pour étudier de nouvelles dimensions des comportements d'inscription et de participation électorales. Ainsi, ces enquêtes permettent désormais d'étudier les personnes inscrites dans une commune différente de celles où elles résident, dites « mal-inscrites » par les auteurs², et les comportements de participation de ces personnes [10]. En 2017, pour la première fois, les agents de la collecte ont relevé si les enquêtés avaient voté par procuration. Cette information nouvelle a permis à Coulmont de montrer que le recours à l'abstention creusait les inégalités sociales : établir une procuration demande des ressources, et notamment de connaître quelqu'un qui pourra porter la procuration (un mandataire) [11]. En 2022, grâce au REU, il ne sera plus nécessaire de faire relever cette information sur les listes d'émargement car elle figurera dans les fichiers. On connaîtra les personnes de l'échantillon mandantes et mandataires, et les dates d'établissements des procurations.

1.4 Diffusion et utilisation des enquêtes de participation électorale

Les enquêtes de participation électorale sont aisément accessibles pour la statistique publique, mais aussi pour la communauté scientifique (chercheurs et étudiants). Des fichiers de production et de recherche et la documentation associée sont disponibles sur le site de l'ADISP (archives de données issues de la statistique publique), service de PROGEDO. Cela concerne les enquêtes de 2002-2004, 2007-2008, 2012 et 2017³. Les données de l'enquête de 2022 y seront également déposées.

Des données plus détaillées sont disponibles au Centre d'accès sécurisé aux données (CASD), après avis favorable du comité du secret.

2. L'échantillon de l'enquête de 2022

Comme les électeurs ont eu jusqu'au 4 mars pour s'inscrire pour participer au 1^{er} tour de l'élection présidentielle de 2022, la constitution de la base de sondage et le tirage de l'échantillon ne peuvent être effectués que durant la seconde partie du mois de mars.

² Les personnes dont la commune de résidence lors de leur recensement est différente de celle de leur inscription sur les listes électorales peuvent avoir déménagé, ou avoir été autorisées effectivement à voter dans une autre commune que celles où elles résident habituellement.

³ Voir par exemple la page consacrée à l'enquête de 2017 : <http://www.progedo-adisp.fr/enquetes/XML/lil.php?lil=lil-1222>.

2.1 Le champ de l'enquête

Le champ de l'enquête est constitué des personnes :

- majeures le 10 avril 2022 ;
- qui résident à Mayotte en janvier 2021 ou dans le reste de la France en janvier 2020 ;
- qui sont inscrites sur les listes électorales principales en métropole ou dans les DOM, avant le 5 mars, afin de pouvoir voter au 1^{er} tour de l'élection présidentielle le 10 avril.

La base de sondage correspondant à ce champ est construite en deux étapes.

Dans un premier temps, l'ensemble des personnes recensées en 2020 (ou 2021), majeures au 10 avril 2022, et qui font partie de l'EDP sont sélectionnées. La probabilité pour un individu d'être dans cette base de sondage est celle d'être présent dans l'EDP, d'avoir été recensé à l'EAR 2020 (ou 2021). Dans l'EDP, des poids (poids EDP, W_{EDP}), inverses de cette probabilité d'inclusion, sont calculés pour que le total pondéré des effectifs de personnes recensées à l'EAR calculé à partir de l'EDP coïncide avec le calcul à partir de l'intégralité de l'EAR, que les personnes soient dans l'EDP ou pas. Ces poids permettent de rendre la base représentative à cette première étape de l'ensemble des personnes résidant en France en janvier 2020 et majeures au 10 avril 2022. Cette représentativité peut être vérifiée, en comparant la pyramide des âges à celle publiée pour le bilan démographique annuel par exemple.

Dans une seconde étape, la base est restreinte aux personnes ayant une inscription active au 10 avril sur les listes électorales principales en France. Cela exclut les personnes de nationalité étrangère, celles qui sont décédées entre janvier 2020 et avril 2022, celles inscrites sur les listes électorales à l'étranger ou dans des communautés d'Outre-mer, celles qui ont été radiées des listes et celles qui ne sont pas inscrites.

Lors de cette seconde étape, il n'y a pas de population à laquelle comparer. Il est impossible de se caler sur la population des personnes inscrites sur les listes, population qui sert au ministère de l'Intérieur pour calculer les taux de participation. En effet, le champ est restreint ici aux personnes résidant en France en janvier 2020 (ou 2021 pour Mayotte). La principale différence provient des personnes inscrites sur les listes électorales en France alors qu'elles résidaient à l'étranger lors du recensement de 2020. Il peut s'agir de personnes qui étaient parties temporairement en janvier 2020, par exemple pour quelques mois, mais sont revenues vivre en France, ou bien de personnes qui résident de façon durable à l'étranger mais ont maintenu leur inscription dans leur commune de résidence antérieure en France.

À partir de la base de sondage ainsi constituée, qui comprend autour de 300 000 personnes, un échantillon de 35 000 électeurs est sélectionné pour lesquels la participation sera relevée sur les listes d'émargement. Le tirage cherche à concilier plusieurs objectifs : avoir un échantillon représentatif sur le champ, limiter la dispersion des poids de tirage, et limiter la dispersion géographique de la collecte.

2.2 Le tirage de l'échantillon

2.2.1 Répartition régionale et sélection des départements enquêtés

On tire environ 35 000 individus inscrits en France. On cherche à avoir un minimum de l'ordre de 1 500 inscrits enquêtés par région, pour pouvoir diffuser des résultats par région. Le surplus par rapport à ce minimum est réparti entre régions au prorata du nombre d'inscrits sur les listes électorales en 2021 (figure 6).

En Guyane et à Mayotte, il y a trop peu de personnes inscrites sur les listes électorales nées un jour EDP et enquêtées à l'EAR 2020 ou 2021 pour pouvoir produire des résultats représentatifs pour ces régions. On ne cherchera donc pas à atteindre le seuil de 1 500 personnes sélectionnées dans ces deux régions, mais un échantillon minimal de 100 personnes dans chacune. Leur intégration est importante pour la représentativité nationale de l'enquête. La taille des échantillons ne permettra toutefois pas la réalisation d'études spécifiques sur ces territoires.

Figure 6. Répartition par région de l'échantillon

Région		Allocation proportionnelle	Allocation retenue	Nombre de départements		
				Total	À exclure	Parmi
01	Guadeloupe	239	autant que disponible , env. 1200	1	-	
02	Martinique	233	autant que disponible , env. 1200	1	-	
03	Guyane	74	100	1	-	
04	La Réunion	500	1500	1	-	
06	Mayotte	65	100	1	-	
11	Ile-de-France	5 473	4700	5 (+ 3)	-	
24	Centre-Val de Loire	1 385	1500	6 + 1*	1	5
27	Bourgogne-Franche-Comté	1 506	1500	8	1	6
28	Normandie	1 818	1600	5 + 1*	1	3
32	Hauts de France	3 220	2 800	5	-	
44	Grand Est	2 933	2 500	10 + 1*	2	5
52	Pays de la Loire	2 101	1 800	5	-	
53	Bretagne	1 896	1 600	4	-	
75	Nouvelle Aquitaine	3 315	2 800	12	3	7
76	Occitanie	3 209	2 700	13	3	10
84	Auvergne-Rhône-Alpes	4 121	3 500	12	3	7
93	Provence-Alpes-Côte d'Azur	2 729	2 300	6	1	3
94	Corse	182	1 500	2	-	
Total		35 000	35 000	101	15	46

* Départements franciliens délégués.

2.2.2 Quinze départements exclus pour limiter les déplacements

Afin de faciliter le travail de collecte et limiter le nombre de déplacements en préfecture, 15 départements ont été exclus de la collecte, comme en 2017. Comme en 2017, le nombre de départements à exclure dans chaque région est défini selon le nombre de départements à la charge de cette région. Cela conduit à exclure des départements dans 8 régions.

En 2017, la sélection des départements était aléatoire dans chaque région, une fois exclus les départements comprenant des établissements régionaux de l'Insee. Il y avait donc implicitement une strate exhaustive, à savoir les départements comprenant des établissements régionaux, même si les probabilités de tirage associée à cette étape n'étaient pas calculées. L'aspect aléatoire avait conduit à exclure des départements comprenant beaucoup d'inscrits dans certaines régions. Par exemple en région PACA, les deux départements exclus représentaient 40% des personnes inscrites de la région.

Pour cette raison, après avoir exploré d'autres solutions (tirage à probabilités proportionnelles au nombre d'électeurs inscrits dans le département, exclusion déterministe des 15 départements comprenant le moins d'inscrits), celle retenue consiste à étendre la strate exhaustive : en plus des départements comportant un établissement régional, cette strate inclut les départements comprenant plus de 500 000 inscrits. Cette nouvelle règle ajoute 13 départements dans la strate exhaustive : ils

sont situés dans une région comprenant au moins 6 départements, ils n'ont pas d'établissement régional et en 2021, plus de 500 000 personnes étaient inscrites sur les listes électorales du département. Il reste alors 46 départements parmi lesquels 15 ont été exclus. Pour l'EPE 2022, il n'y aura pas de collecte dans ces 15 départements : Allier, Ardennes, Cher, Deux-Sèvres, Dordogne, Hautes-Alpes, Haute-Loire, Lot, Lot-et-Garonne, Lozère, Manche, Meuse, Saône-et-Loire, Savoie, Tarn.

2.2.3 Commune de recensement, commune d'inscription et strates de tirage

Pour faciliter la collecte, celle-ci a lieu uniquement dans les listes d'émargement des communes recensées en 2020. Cela signifie que seules les personnes inscrites dans une commune enquêtée à l'EAR 2020 peuvent être sélectionnées pour faire partie de l'échantillon⁴. Si elles sont inscrites dans leur commune de recensement, alors par construction elles peuvent être sélectionnées. Ce n'est pas systématique pour celles qui sont inscrites dans une commune différente de celle de leur recensement. En effet, si leur commune d'inscription est une petite commune au sens du recensement (moins de 10 000 habitants), elle a une chance sur 5 d'avoir été recensée en 2020. C'est pour cela que nous avons repris le découpage de 2017 en six strates, selon la taille de la commune de recensement et de celle d'inscription, afin de mieux représenter les personnes inscrites dans une commune différente de celle où elles ont été recensées (**figure 7**).

Le poids de tirage à cette étape est $w_i = w_{EDP} \times w_{DEP} \times w_{INS_EAR}$.

w_{INS_EAR} est l'inverse de la probabilité que la commune d'inscription ait été recensée en 2020, il vaut 5 pour les individus des strates 3 et 4 et il vaut 1 sinon.

Figure 7. Répartition par strate de l'échantillon

Strates		En %	
		2017	2022*
1	Inscrits dans leur commune de résidence qui est une petite commune	47	47
2	Inscrits dans leur commune de résidence qui est une grande commune	36	36
3	Inscrits dans une petite commune, qui résident dans une autre petite commune	6	5
4	Inscrits dans une petite commune, qui résident dans une grande commune	4	5
5	Inscrits dans une grande commune, qui résident dans une petite commune	2	2
6	Inscrits dans une grande commune, qui résident dans une autre grande commune	5	5
Total		100	100

* Provisoire décembre 2021.

La répartition par strate est extrêmement similaire à celle de 2017. On calcule ensuite une allocation théorique par strate et par région, en fonction de la répartition pondérée dans la base de sondage.

⁴ En Guadeloupe, à la Martinique, à La Réunion et en Corse, cette restriction n'est pas appliquée : toutes les personnes recensées, donc qui résident dans une commune recensée en 2020, sont enquêtées, même si elles sont inscrites dans une commune, différente de leur lieu de résidence, qui n'a pas été recensée.

2.2.4 Tirage de l'échantillon dans chaque strate

- **Les personnes inscrites en grande commune (strates 2, 5 et 6)**

Toutes les grandes communes sont enquêtées, donc les personnes inscrites dans une grande commune sont toutes susceptibles d'être échantillonnées (strates 2, 5, 6).

Parmi elles, les personnes résidant dans des grandes adresses ou des adresses nouvelles, ainsi que les personnes de la strate 5 (inscrites dans une grande commune, recensées dans une petite commune) ont plus de chances que les autres d'avoir été enquêtées (environ trois fois plus de chances). Elles ont donc des poids initiaux trois fois inférieurs. Afin d'uniformiser les poids, on sélectionne aléatoirement une personne sur trois dans ces catégories et on multiplie le poids des personnes sélectionnées par trois. Passée cette étape, la sélection des individus se fait de façon aléatoire de façon à atteindre les allocations en grande commune définies dans chaque région.

- **Les personnes inscrites en petite commune (strates 1, 3 et 4)**

Pour la strate 1, afin de limiter le coût de collecte, on veille à limiter le nombre de communes enquêtées. On souhaite tirer 8 individus par commune environ, comme en 2012 et en 2017, pour ne pas avoir trop d'effet de grappe sans trop augmenter le nombre des communes à enquêter. On sélectionne donc environ 8 fois moins de petites communes dans chaque région que l'allocation visée pour la strate 1. Le tirage de ces petites communes se fait région par région, par tirage systématique à probabilité inégale, proportionnellement au nombre d'inscrits, sur une base triée par canton pour améliorer la dispersion géographique des communes sélectionnées. Une fois sélectionnées les petites communes et le nombre d'individus à sélectionner dans chacune déterminé, le tirage des individus est fait par tirage systématique à probabilités égales sur base triée par âge, diplôme, indicatrice immigré ou non. Ce tri par variable permet de conserver de manière plus sûre qu'un simple tirage aléatoire la représentativité souhaitée sur ces variables qui jouent sur la participation au vote. La méthode assure que tous les individus de la strate 1 d'une même région ont la même probabilité d'être sélectionnés à cette étape, proche d'une chance sur 10 dans toutes les régions, sauf en Guadeloupe, en Martinique et en Corse où la probabilité vaut 1. Les poids de tirage sont multipliés par l'inverse de cette probabilité régionale, noté w_{INDCOM} .

Pour la strate 3, les personnes retenues sont celles qui sont inscrites dans une commune sélectionnée au titre de la strate 1, les individus sont ensuite tirés au hasard en fonction de leur probabilité d'avoir été sélectionnés à cette étape. Cela permet de réduire la dispersion des poids, particulièrement importante au sein de cette strate. En effet, w_{INDCOM} est l'inverse de la probabilité que l'individu ait été sélectionné, c'est-à-dire la probabilité que l'individu ait été sélectionné dans sa commune (égale à 1), multipliée par la probabilité que la commune ait été sélectionnée pour la strate 1. w_{INDCOM} est donc variable et liée au nombre d'inscrits en strate 1.

Pour la strate 4, l'échantillon disponible est inférieur à l'allocation théorique. Toutes les personnes sont donc retenues, dès lors que leur commune d'inscription ait été recensée. Cela augmente par conséquent le nombre de communes à enquêter.

2.2.5 Un échantillon de 35 000 électeurs

Les poids de tirage de cet échantillon de 35 000 personnes seront assez dispersés. Cela vient de la complexité du plan de sondage, héritée en partie de celle de l'enquête de recensement.

À l'issue de ces étapes, les poids de tirage valent $w_i = w_{EDP} \times w_{DEP} \times w_{INS_EAR} \times w_{INDCOM}$ en petites communes et $w_i = w_{EDP} \times w_{DEP} \times w_{INS_EAR} \times w_{IND}$ en grandes communes.

Dans certaines régions, notamment dans les DROM, toutes les personnes de la base de sondage sont échantillonnées : leur poids de tirage est l'inverse de la probabilité d'être né un jour EDP et d'avoir été recensé en 2020, de l'ordre de $14\% \times 4,3\%$, soit un poids en moyenne autour de 170. Dans d'autres régions la probabilité d'être sélectionné est faible, surtout pour une personne inscrite dans une petite commune différente de sa commune d'inscription, vivant dans un département de taille moyenne. La probabilité de tirage est plus faible car la probabilité d'être née un jour EDP et d'avoir été recensée est multipliée par la probabilité que le département de résidence soit enquêté, que la commune d'inscription soit recensée (1/5) et qu'elle ait été sélectionnée pour être enquêtée à l'EPE2022. Les poids de tirage de ces individus seront donc beaucoup plus élevés.

La composition de l'échantillon est comparée à celle de la base de sondage (âge, sexe, diplôme, taille de la commune). Elles sont proches en général.

Toutefois, la population enquêtée peut être un peu plus distincte car une partie des personnes échantillonnées ne seront pas retrouvées dans les listes et ces personnes auront peut-être des caractéristiques distinctes. Elles sont assimilées à des non-répondants. En 2017, sur 44 998 individus retenus dans l'échantillon, 909 n'ont pas été retrouvés au moins une fois sur les listes électorales lors de la collecte pour l'élection présidentielle, soit 2,1 %. Il s'agissait d'individus non retrouvés sur les listes où ils auraient dû figurer d'après les informations de la base de tirage, ou de personnes non retrouvées parce que la liste le concernant n'a pas pu être consultée en préfecture (liste manquante). La tranche d'âge 18-29 ans était la moins concernée (1,8 % de non repérés). Entre 25 ans et 74 ans, le taux de non retrouvés oscillait autour de 2 %. Il augmentait ensuite avec l'âge : 3,8 % de non retrouvés à 85 ans ou plus.

En 2022, la structure par âge des personnes non retrouvées devrait être similaire. On peut en revanche espérer une diminution de la proportion de personnes non retrouvées par rapport à 2017, puisque la mise en place du répertoire électoral unique, duquel est directement extrait les listes électorales, devrait permettre une très bonne cohérence des informations dans la base de sondage et sur les listes électorales des mairies.

Le calage sera effectué en utilisant les marges calculées sur la base de sondage définie au 2.1. Les modalités exactes restent à définir, ainsi que les variables prises en compte (sexe, âge, diplôme ou catégorie sociale, caractéristiques de la commune de résidence). Une attention particulière devra être portée à la représentativité régionale puisque les résultats régionaux seront diffusés.

Le poids initial pour ce calage sera l'inverse de la probabilité de tirage.

3. Le déroulement de l'opération

3.1 Calendrier

En 2022, les deux tours de l'élection présidentielle auront lieu les 10 avril et 24 avril, et les élections législatives auront lieu les 12 et 19 juin. La préparation de l'enquête a commencé fin 2021, l'opération sera finalisée fin 2022 avec la mise à disposition du fichier des données statistiques documenté : le calendrier est donc assez resserré (**figure 8**).

Figure 8 : Calendrier de l'enquête EPE 2022

Mars-avril	Extraction de la base par le pôle RFD Tirage définitif de l'échantillon des communes et individus Livraison des bordereaux de collecte et prise en main par les agents Contact avec les préfetures – communication dans la presse régionale Visio-conférences de formation
Avril-mai	<i>Élection présidentielle (1^{er} tour : 10 avril, 2^e tour : 24 avril)</i> Collecte « Présidentielle » en préfecture (jusqu'au 4 mai)
Juin	<i>Élections législatives (1^{er} tour : 12 juin, 2^e tour : 19 juin)</i> Collecte « Législatives » en préfecture (jusqu'au 29 juin, ou 22 s'il n'y a qu'un tour)
Juillet-août	Questionnaire de bilan de collecte Constitution du fichier définitif, pondéré et documenté Fichier mis à disposition des DR
Septembre - novembre	Note sur le redressement et la pondération Publications nationales et régionales Diffusion du fichier statistique (Adisp et CASD)

3.2 La collecte

La collecte se déroule principalement dans les dix jours qui suivent chaque second tour de scrutin. Il peut y avoir des exceptions qui nécessitent de collecter l'information dès le 1^{er} tour, si la commune a des listes d'émargement séparées ou, dans le cas des élections législatives, s'il n'y a qu'un tour dans la circonscription.

L'organisation de la collecte se fait avec des correspondants EPE, dans chaque région ou établissement. Des agents des directions régionales se rendent dans chaque préfecture pour consulter les listes d'émargement des bureaux de vote dans lesquels sont inscrits les électeurs. Pour relever l'information, ils utilisent des bordereaux de collecte. Un bordereau est un fichier Calc départemental, qui contient un onglet par commune. Dans chaque onglet figure la liste des électeurs dont la participation doit être relevée (**figure 9**).

Pour l'enquête 2022, la collecte va être grandement simplifiée grâce au REU. Dans le fichier électoral utilisé en 2017, le numéro d'inscription et surtout le bureau de vote de l'électeur n'étaient pas connus au moment de tirer l'échantillon. Or il y a environ un bureau de vote pour 1 000 électeurs, et donc il y a de nombreux bureaux de vote dans les grandes villes. Il serait bien trop fastidieux de consulter la liste d'émargement de chaque bureau de vote de la ville. Dans l'exemple ci-dessous, on voit que les 18 électeurs échantillonnés à Valenciennes sont dans 14 bureaux de vote distincts. En 2017, une opération préalable avait été nécessaire pour relever le bureau de vote, en contactant les mairies. Grâce au REU, cette information est directement disponible dans les bordereaux envoyés aux directions régionales.

Deuxième simplification importante, en 2017 les agents de directions régionales devaient relever si l'électeur avait voté lui-même ou par procuration. Relever cette information n'est pas évident car la notation sur la liste n'est pas normalisée d'une commune à l'autre, ni même au sein d'une commune. Parfois il est noté VPP (pour vote par procuration), parfois une étiquette est collée. On note que l'électeur est mandant (il a demandé à quelqu'un de voter pour lui), mais aussi parfois sur la liste, on mentionne qu'il est mandataire, c'est-à-dire qu'il a un mandat pour voter à la place de quelqu'un d'autre. Cette grande variété avait complexifié le travail de relevé en préfecture. Or désormais l'information figure dans le REU et une extraction sera faite au jour de chaque tour de scrutin, afin de connaître les personnes de l'échantillon qui ont des procurations actives en tant que mandant et en tant que mandataire au moment de ce tour d'élection. Les agents chargés de la collecte doivent donc uniquement relever si la personne a voté (il y a une signature), ou non.

Figure 9 : Bordereau du département du Nord (59), onglet de Valenciennes (données fictives)

EPE - ÉLECTION PRÉSIDENTIELLE 2022										
3	Département : Nord (59)			Préfecture : Lille		Retour à la liste des communes				
5	COMMUNE : Valenciennes (59606)			Nb électeurs total : 18		Trier les électeurs par : <input type="button" value="Bureau et nom"/> <input type="button" value="Bureau et n° d'ins."/> <input type="button" value="Nom de famille"/>				
6				Nb bureaux de vote : 14						
8	Etat de la collecte : Commencée			Nb électeurs renseignés : 25						
10	N° d'échantillon	Etat civil	Prénoms	Date de naissance	Bureau de vote	N° d'inscription	Participation		Remarques	Si autre, précisez :
11		Nom					Tour n°1	Tour n°2		
12	1485	BOU	Val	1999	1	1396	1	1		
13	1486	DAL	Bru	1961	10	449	2	2		
14	1487	SPE	Chr	1948	12	1033	3	3		
15	1488	BON	Lud	1980	13	646	1	1		
16	1489	BOU	Thi	1984	14	128	1	2		
17	1490	DEL	Aga	1997	16	1455	1	1		
18	1491	DES	Val	1971	18	442	2	2		
19	1492	DUM	Ghi	1945	18	1175	2	2		
20	1493	GAI	Bap	2001	19	472	2	2		
21	1494	GIL	Ber	1958	2	873	1			
22	1495	HEN	Ber	1955	2	47	1			
23	1496	SIE	Sté	1977	21	547	1			
24	1497	VAS	Ana	1984	23	304	1			
25	1498	DEL	Edi	1950	3	374	1			
26	1499	FOU	Dav	1976	4	386	1	1		

Bibliographie

- [1] Morin J. La participation électorale dans les grandes villes aux élections municipales de mars 1983, *Économie et statistique*, N°165, avril 1984.
- [2] Durier S., L'échantillon démographique permanent a 50 ans : retours sur un dispositif statistique original, Présentation aux Journées de méthodologie statistique, Paris, juin 2018.
- [3] Robert-Bobée I., Gualbert N., L'échantillon démographique permanent : en 50 ans, l'EDP a bien grandi !, *Courrier des statistiques* n°6, juillet 2021.
- [4] Braconnier, C., Coulmont, B. & Dormagen, J. (2017). Toujours pas de chrysanthèmes pour les variables lourdes de la participation électorale: Chute de la participation et augmentation des inégalités électorales au printemps 2017. *Revue française de science politique*, 67, 1023-1040.
- [5] Héran F., Rouault D., La double élection de 1995 : exclusion sociale et stratégie d'abstention, *Insee première*, n° 414, novembre 1995.
- [6] Buisson G. et Penant S., Participation aux élections présidentielle et législatives de 2017 : neuf inscrits sur dix ont voté à au moins un tour de scrutin, *Insee première*, n° 1670, octobre 2017.
- [7] Buisson G. et Penant S., Élections présidentielles et législatives de 2002 à 2017 : une participation atypique en 2017, *Insee première*, n° 1671, octobre 2017.
- [8] Besnard S., Delhomme I. (2017). Élections 2017 - Voter, une pratique moins systématique qu'avant. Sonia Besnard, Isabelle Delhomme, *Insee Flash Pays de la Loire*, n°73.
- [9] Tiberj V., Entendre les voix de l'abstention. *Hommes & libertés*, Ligue des droits de l'Homme, 2020, Manifestation(s), pp.12-14.
- [10] Braconnier, C., Dormagen, J., Gabalda, G. & Niel, X. (2016). Sociologie de la mal-inscription et de ses conséquences sur la participation électorale. *Revue française de sociologie*, 57, 17-44.
- [11] Coulmont, B. (2020). *In absentia*: Le vote par procuration, une participation électorale à distance ?. *Revue française de science politique*, 70, 469-488.
- [12] Durier S., Touré G., Élections de 2017 : 6,5 % des citoyens ont fait une démarche volontaire pour s'inscrire, *Insee Focus*, n°80, mars 2017.