

Séminaire Fabrique des données quantitatives

(Dé)construire un indicateur statistique
L'exemple du taux de chômage et du solde migratoire

L'ECOLE
DES HAUTES
ETUDES EN
SCIENCES
SOCIALES

PSL 
RESEARCH UNIVERSITY PARIS

Martin CHEVALIER (Insee)

Année universitaire 2019-2020

Objectifs de cette séance

1. Entrouvrir le couvercle de la boîte noire, la **porte de la cuisine** : comment sont produits en pratique certains des indicateurs statistiques les plus commentés de l'Insee ?
2. Fournir des **points de repères** et alimenter une **culture générale de la statistique publique**, à travers certains de ses dispositifs les plus structurants.
3. Souligner le **caractère conventionnel des indicateurs**, à toujours garder à l'esprit au moment de l'interprétation.

Partir des publications pour remonter aux données

1. Connaître et savoir lire la publication où apparaît l'indicateur
2. Enjeux autour de la définition} de l'indicateur
3. La source : présentation de l'enquête servant à calculer l'indicateur
4. La construction du chiffre (1) : questionnaire et questionnement
5. La construction du chiffre (2) : sondage et modélisation

Deux thématiques, deux enquêtes :

- ▶ le **taux de chômage** : indicateur clé pour analyse la conjoncture du marché du travail → **enquête Emploi en continu** (EEC) ;
- ▶ le **solde migratoire** : résultante du bilan démographique particulièrement commentée → **Recensement de la population** (RP).

Quelques mots de présentation de l'Insee

Insee : Institut national de la statistique et des études économiques

- ▶ producteur majeur de données socio-économiques : PIB, taux de chômage, démographie, enquêtes sociales, etc. ;
- ▶ acteur central du Système statistique public (SSP) qui coordonne les Services statistiques ministériels (SSM) : Dares, Drees, etc.
- ▶ 5 400 salariés, dont 2 200 fonctionnaires de catégorie B et 1 600 de catégorie A ou A+ (au 31/12/2018).

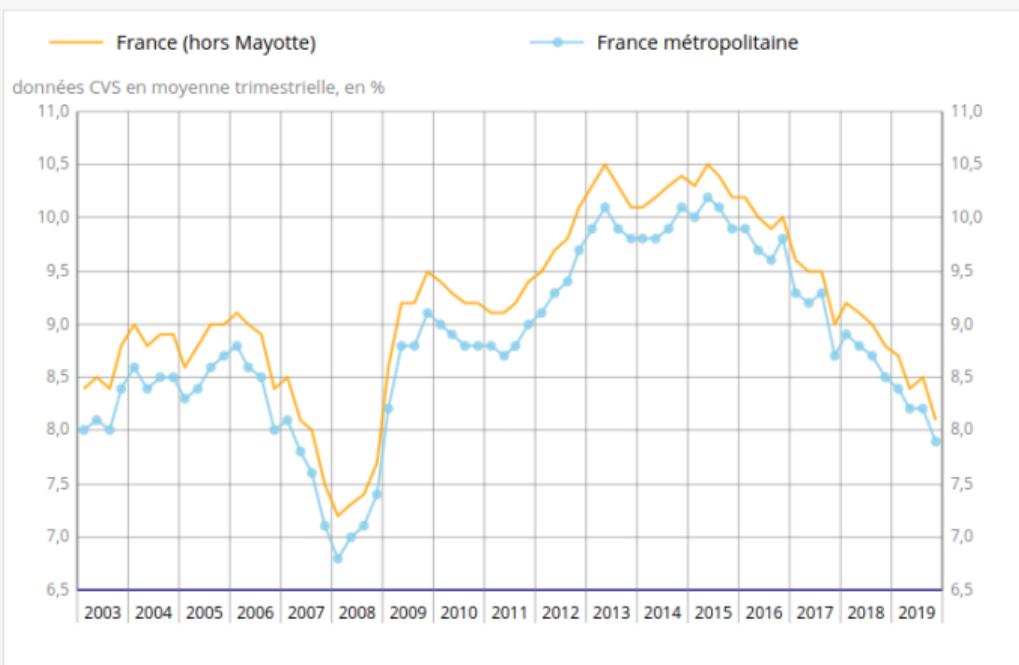
Travailler à l'Insee

- ▶ implantation sur l'ensemble du territoire (Paris et capitales de région) ;
- ▶ de nombreux concours pour des postes à profil statistique : contrôleur (B), attaché (A), administrateur (A+) ;
- ▶ formation rémunérée de un an (B, à Libourne) ou deux ans (A à Rennes, A+ à Saclay) ;
- ▶ environnement et salaires attractifs : 2 400 € net pour un attaché en début de carrière, 3 200 € pour un administrateur.



Le taux de chômage et l'enquête Emploi en continu

Taux de chômage au sens du BIT



Estimation à +/- 0,3 point près du niveau du taux de chômage et de son évolution d'un trimestre à l'autre.

Champ : population des ménages, personnes de 15 ans ou plus.

Source : Insee, enquête Emploi.

« Informations rapides » : publication récurrente (ici trimestrielle), à structure fixe, soumise à un calendrier de publication strict.

Structure Taux de chômage BIT, halo, taux d'emploi BIT, temps partiel et sous-emploi, taux d'activité. → beaucoup de concepts à maîtriser pour interpréter correctement la publication.

Autres éléments notables :

- ▶ « données CVS » : données corrigées des variations saisonnières, *i.e.* lissées de sorte que la série ne soit pas affectée par des phénomènes purement saisonniers ;
- ▶ « France (hors Mayotte) » : Mayotte est un département français depuis le 31 mars 2011. Il n'est cependant pas encore intégré dans l'enquête Emploi en continu servant à calculer le taux de chômage trimestriel ;
- ▶ « population des ménages » : population résidant en logement ordinaire, *i.e.* hors communautés (résidences étudiantes, foyers de travailleurs, maisons de retraite, etc. : 1,6 M de personnes environ) et hors sans-abri, habitations mobiles, bateliers.

Enjeux de définition : Le taux de chômage BIT

Du point de vue du marché du travail, une population peut être décomposée schématiquement en plusieurs ensembles :

- ▶ les **actifs**, qui occupent un emploi ou en recherchent un, se distinguent des **inactifs** ;
- ▶ parmi les actifs, les **actifs occupés** occupent un emploi et se distinguent des **chômeurs** qui en recherchent un.

La frontière entre actifs occupés, chômeurs et inactifs est précisément définie au niveau international par le **Bureau international du travail** (BIT). Est considéré comme au chômage toute personne de **15 ans ou plus** :

1. qui n'a **pas travaillé**, ne serait-ce qu'une heure, au cours de la semaine de référence ;
2. qui est **disponible** pour travailler dans les deux semaines ;
3. qui a entrepris des **démarches actives** de recherche d'emploi dans le mois précédent (ou a trouvé un emploi qui commence dans les trois mois).

Le taux de chômage BIT d'un trimestre est le **rapport entre le nombre de chômeurs et le nombre d'actifs** en moyenne au cours de ce trimestre.

Enjeux de définition : Taux d'emploi et taux d'activité

Le taux de chômage constitue la **mesure de référence** pour analyser la conjoncture du marché du travail. D'autres peuvent néanmoins être mobilisées pour **affiner ou nuancer l'analyse** :

- ▶ **taux d'emploi** : part que représentent les actifs occupés dans l'ensemble de la population d'une classe d'âge donnée (exemple : taux d'emploi des 55-64 ans).
Le taux d'emploi est **indépendant de la partition entre actifs et inactifs**. Contrairement au taux de chômage, il n'est donc pas affecté par les **transitions du chômage à l'inactivité** (inaptitude, retraite anticipée, etc.).
- ▶ **taux d'activité** : part que représentent les actifs (occupés ou chômeurs) dans l'ensemble de la population d'une classe d'âge donnée (exemple : taux d'activité des 25-30 ans).

Enjeux de définition : Halo et sous-emploi

Le taux de chômage et le taux d'emploi constituent les **mesures de référence**, notamment au niveau international (BIT), mais ne permettent pas toujours de **capter toute la complexité des situations vis-à-vis de l'emploi**. Des mesures complémentaires ont ainsi été introduites :

- ▶ **halo du chômage** : les individus dans le halo du chômage sont inactifs mais présentent **deux des trois critères définissant le chômage BIT**. En pratique, cela signifie :
 - ▶ soit qu'ils n'étaient pas disponibles pour travailler dans les deux semaines (formation, etc.) ;
 - ▶ soit qu'ils n'ont pas entrepris de recherche active d'emploi (chômeurs découragés, etc.).
- ▶ **sous-emploi** : les individus en situation de sous-emploi sont actifs occupés mais ont travaillé la semaine de référence **moins que souhaité**, en règle générale du fait d'un **temps partiel subi**.

Enjeux de définition : Taux de chômage et DEFM (1)

La statistique publique (Pôle emploi et Dares) produit une autre mesure majeure de la conjoncture du marché du travail : le **nombre de Demandeurs d'emploi en fin de mois** (DEFM).

Il s'agit d'une mesure construite à partir des **données de gestion de Pôle emploi**, en l'espèce le nombre d'individus inscrits dans différentes catégories :

- ▶ catégorie A : Personne sans emploi, à la recherche d'un emploi quel que soit le type de contrat (CDI, CDD, à temps plein, à temps partiel, temporaire ou saisonnier) ;
- ▶ catégorie B et C : Personne ayant exercé une activité réduite ;
- ▶ catégorie D : Personne sans emploi, qui n'est pas immédiatement disponible ;
- ▶ catégorie E : Personne pourvue d'un emploi.

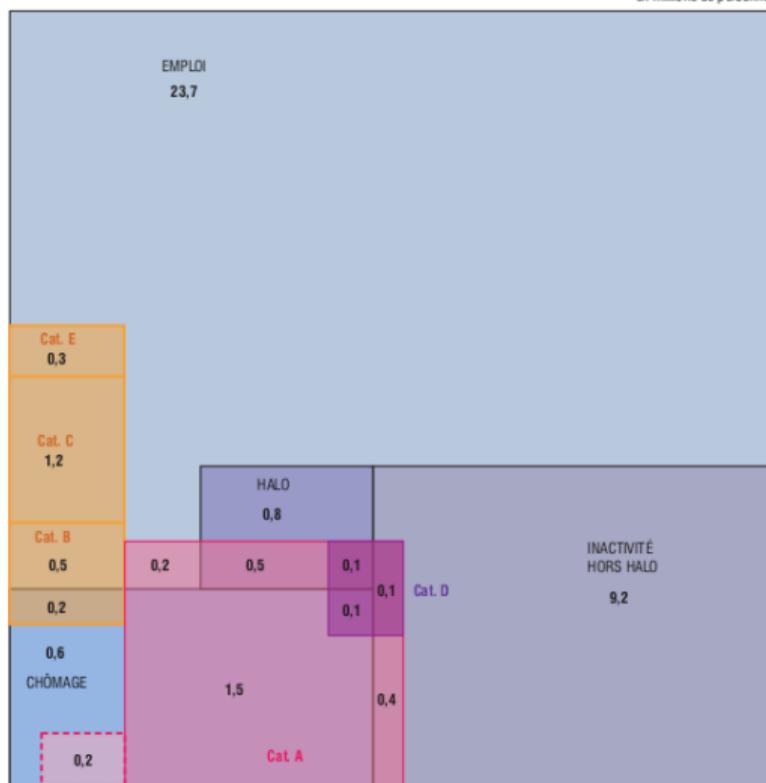
Les concepts, sources et méthodes

diffèrent sensiblement de la mesure du chômage BIT.

Enjeux de définition : Taux de chômage et DEFM (2)

2. Superposition des statuts BIT et des catégories de Pôle emploi en 2017

en millions de personnes

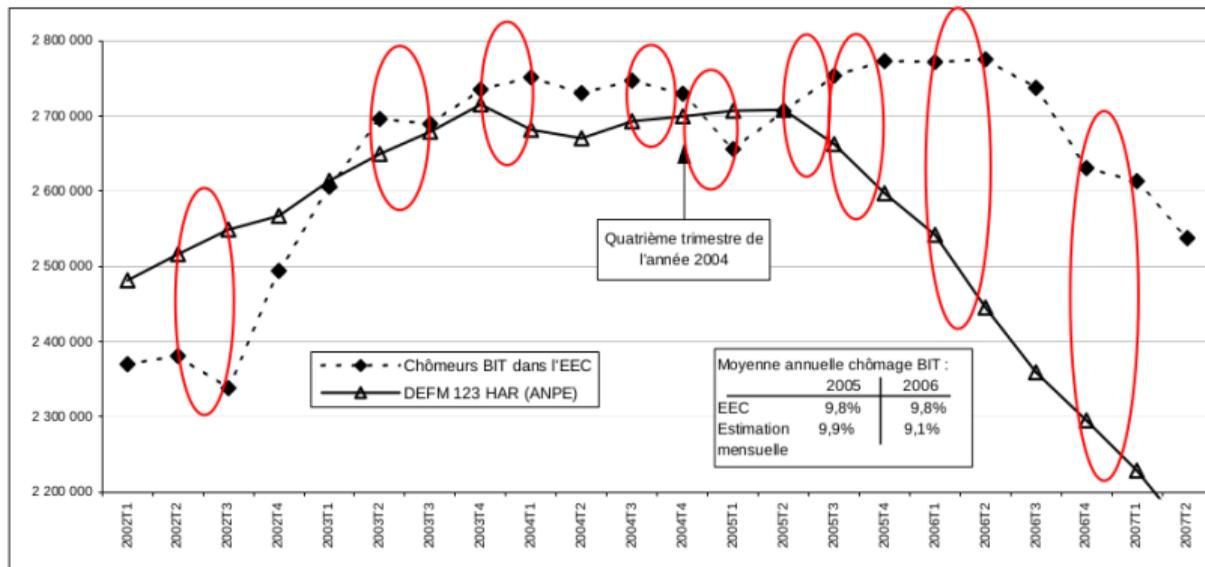


Le taux de chômage et l'enquête Emploi en continu

Enjeux de définition : Taux de chômage et DEFM (3)

Des divergences importantes ont pu avoir lieu par le passé entre inscrits à Pôle emploi et chômage BIT, notamment en 2007.

Figure 1 : évolution trimestrielle des chômeurs BIT selon l'enquête emploi et des DEFM 123 HAR (ANPE) de 2002 à 2007



Source : Insee

Le taux de chômage et l'enquête Emploi en continu

La source : l'enquête Emploi en continu

L'enquête Emploi en continu constitue, avec le recensement de la population, **une des opérations statistiques les plus importantes** menées par l'Insee :

- ▶ 90 000 logements tirés au sort, 110 000 personnes de 15 ans et plus chaque trimestre ;
- ▶ enquête en continu depuis 2003 : près de 7 000 logements enquêtés pour chaque semaine de l'année (« semaine de référence ») ;
- ▶ enquête **en panel** : les logements sont interrogés six trimestres de suite.

Elle comporte un grand nombre de « modules » permettant de **décrire la situation des personnes** sur le marché du travail mais aussi leurs **conditions de travail** :

- ▶ module A (ou module BIT) : position sur le marché du travail ;
- ▶ module B : profession (PCS), type de contrat, horaires, temps de travail souhaité, conditions de travail, revenus ;
- ▶ module D : formation initiale, formation continue ;
- ▶ module G : origine géographique et sociale ;
- ▶ module H : santé (mini-module européen sur la santé perçue).

Le taux de chômage et l'enquête Emploi en continu

Opérationnalisation de la définition BIT

La détermination de la position sur le marché du travail au sens du BIT mobilise **l'ensemble des questions du module A** (plus d'une trentaine). Celles-ci sont regroupées en plusieurs ensembles :

1. Exercice d'une activité professionnelle effective :
 - ▶ importance de la question sur l'exercice d'une activité ne serait-ce qu'une heure ;
 - ▶ prise en compte des situations d'emploi rémunéré sans travail la semaine de référence.
2. Recherche effective d'un travail et méthodes de recherche d'emploi :
 - ▶ un grand nombre de modalités de recherche d'emploi sont listées (yc relations personnelles, réseaux sociaux, salon ou bourse d'emploi, etc.) ;
 - ▶ prise en compte des personnes souhaitant s'installer à leur compte (recherches de locaux ou de financement, etc.)
3. Disponibilité pour travailler dans les deux semaines (y compris quand un autre emploi est prévu plus de trois mois après la semaine de référence).

Le taux de chômage et l'enquête Emploi en continu

Modes de collecte et rôle de l'enquêteur

Comme de nombreuses enquêtes de l'Insee, l'enquête Emploi en continu est **longue et complexe** :

- ▶ mesure **très précise** des concepts nécessaires au chômage BIT ;
- ▶ prise en compte de situations **très différentes** → **nombreux filtres et questions conditionnelles**.

Par ailleurs, l'enquête Emploi en continu est une **enquête multi-mode** :

- ▶ première interrogation en face-à-face ;
- ▶ réinterrogations suivantes par téléphone (bientôt internet).

Dans ce contexte, la bonne réalisation de l'enquête dépend de façon cruciale des **outils et des compétences à disposition des enquêteurs** :

- ▶ déjà, **outils informatiques** : questionnaires CAPI et CATI adaptés, notamment au contexte de réinterrogation ;
- ▶ surtout, **formation et compétences spécifiques** : respect des consignes et des formulations, mais aussi « **traduction** » pour être compris par les enquêtés.

Refonte du questionnaire de l'enquête Emploi en continu en 2013 : une des conséquences de la polémique sur les chiffres du chômage de 2007.

Objectifs de la refonte du questionnaire :

- ▶ faire que les questions « passent mieux » sans reformulation ;
- ▶ notamment dans le contexte des réinterrogations téléphoniques (temps plus contraint).

En pratique pour le module A : reformulations, changement de position, ajouts de questions.

Conséquences -145 000 personnes au chômage (-0,5 points), +38 000 en emploi et + 107 000 inactives dans le halo.

Référence [Présentation au CNIS](#)

Le taux de chômage et l'enquête Emploi en continu

De l'importance de la formulation et de l'ordre des questions (2)

Ancien questionnaire	Questionnaire rénové
EMPLOI	
	...(emploi informel) avez-vous effectué une ou plusieurs activités, déclarée(s) ou non, pour faire face à vos dépenses ?
CHÔMAGE	
<Recherche d'emploi au cours du dernier mois> ...étiez-vous à la recherche d'un emploi même à temps partiel, même occasionnel ?	<Souhait de travailler> <Recherche d'emploi au cours du dernier mois> ... avez-vous recherché un emploi ?
<Démarches actives de recherche d'emploi> Questions sur les démarches de recherche dont : Avez-vous été contacté par Pôle emploi ...?	<Démarches actives de recherche d'emploi> Questions sur les démarches de recherche suppression de la modalité contact par Pôle emploi en 2013 (Eurostat) autres modalités rajoutées (mobilisation de réseaux sociaux)
<Disponibilité dans les 15 jours> Si on vous proposait un emploi, seriez-vous prêt à travailler dans un délai inférieur à deux semaines ?	<Disponibilité dans les 15 jours> Si vous trouviez un travail qui vous convienne, seriez-vous disponible pour le commencer dans les deux semaines ?

De l'importance de la formulation et de l'ordre des questions (3)

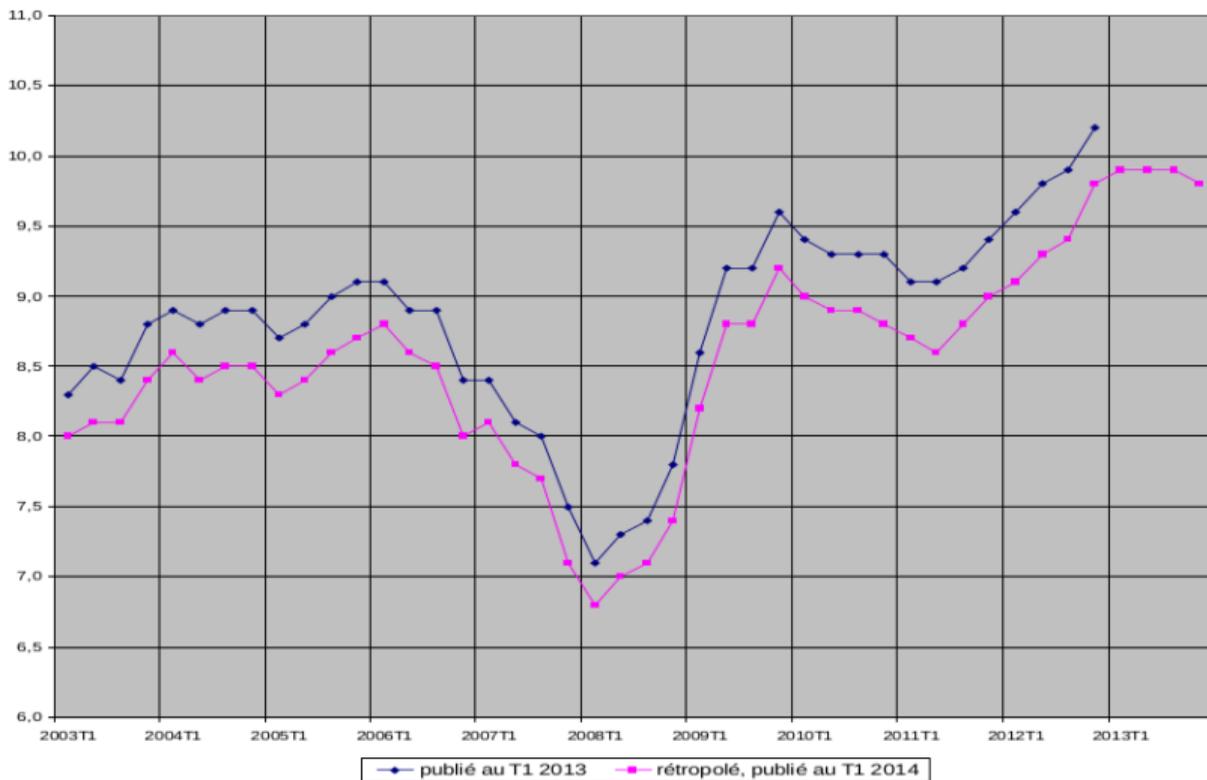
Plusieurs éléments expliquent l'évolution du nombre de chômeurs avec la refonte du questionnaire :

- ▶ (-) reformulation de la question sur la recherche d'emploi, moins de confusion avec le souhait de travailler ;
- ▶ (-) suppression de la modalité "Pôle emploi" ;
- ▶ (-) ajout de la question sur le travail informel ;
- ▶ (+) nouvelles modalités de recherche d'emploi (internet) ;
- ▶ (+) reformulation de la question sur la disponibilité.

Conséquences pratiques

- ▶ **perturbation de la publication** du taux de chômage 2013 T1 ;
- ▶ **rétopolation** des séries historiques et des fichiers détails.

Taux de chômage cvs, France métropolitaine, en %



Impact du sondage sur la précision des résultats (1)

Un élément du graphique **laissé en suspens** : « Estimation à +/- 0,3 point près du niveau du taux de chômage et de son évolution d'un trimestre à l'autre. »

L'enquête Emploi n'étant **pas exhaustive**, tous les indicateurs construits à partir de cette source comportent une **imprécision liée au tirage au sort**.

Des résultats de théorie des sondages **garantissent que** :

- ▶ le fait de s'appuyer sur un sondage ne conduit **pas à sur- ou sous-estimer systématiquement l'indicateur visé** : on parle d'estimateur **sans biais**);
- ▶ l'imprécision associée au sondage est **susceptible d'être estimée** : cette **estimation de variance** est en pratique complexe.

Cas du sondage aléatoire simple

Le sondage aléatoire simple est le plan de sondage **le plus simple qui soit**, qui correspond en pratique par exemple au **tirage du loto**.

Si on cherche à estimer une proportion à partir d'un échantillon d'individus sélectionnés par sondage aléatoire simple, alors l'**imprécision de l'estimateur de cette proportion** \hat{p} est :

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1 - \hat{p})}{n - 1}}$$

avec n le nombre d'observations de l'échantillon : **plus n est grand, plus l'incertitude est petite**.

$\hat{\sigma} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n - 1}}$ est appelée l'**écart-type de \hat{p}** . Une mesure alternative est l'**imprécision relative** $\hat{\delta} = \frac{2\hat{\sigma}}{\hat{p}}$.

Impact du sondage sur la précision des résultats (3)

Cas du sondage aléatoire simple

Relation entre valeur anticipée de la proportion, imprécision relative et taille de l'échantillon nécessaire.

	0,05	0,10	0,20	0,30	0,40	0,50
1 %	760000	360000	160000	93333	60000	40000
2 %	190000	90000	40000	23333	15000	10000
3 %	84444	40000	17778	10370	6667	4444
4 %	47500	22500	10000	5833	3750	2500
5 %	30400	14400	6400	3733	2400	1600
10 %	7600	3600	1600	933	600	400

Lecture Pour une proportion anticipée de 10 % et une précision relative de 2 % (ie un intervalle de confiance de $\pm 0,2$ points de pourcentage), la taille d'échantillon nécessaire est de 90 000.

Impact du sondage sur la précision des résultats (4)

Plusieurs éléments dans le plan de sondage de l'EEC **diffèrent sensiblement du sondage aléatoire simple** :

- ▶ **concentration de la collecte** : les logements ont été tirés « en grappes » (*i.e.* en paquets compacts sur les territoires) de façon à faciliter le travail des enquêteurs (en temps très contraint : 2 semaines et 2 jours) ;
- ▶ **structure de panel** : la réinterrogation des logements six trimestres consécutifs allège le protocole et permet une **meilleure mesure des évolutions** ;
- ▶ **utilisation de techniques de sondage sophistiquées** : des informations auxiliaires sur les logements sont disponibles dans la base de sondage (fichiers fiscaux). En pratique, on fait en sorte que **tous les profils de ménages soient représentés** (en restant dans un cadre aléatoire).

Exemple Stratification par région : plutôt que de tirer au niveau national, on tire région par région. De la sorte, on **contrôle totalement le nombre de logements tirés par région** tout en restant dans un cadre aléatoire.

Le taux de chômage et l'enquête Emploi en continu

Le rôle des redressements : la correction de la non-réponse

Les enquêtes auprès des ménages de l'Insee font face à une **non-réponse de plus en plus importante**.

L'enquête Emploi en continu est une enquête prioritaire sur laquelle un important effort de relance est concentré, néanmoins **le taux de réponse est souvent supérieur à 30 %**.

Cette non-réponse ne serait pas problématique **si les caractéristiques des non-répondants ne différaient pas de celles des répondants** ; or c'est vraisemblablement le cas.

Une **correction de la non-réponse** est mise en œuvre de façon à redresser l'échantillon en exploitant de nombreuses variables (pyramide des âges régionale, revenus, etc.).

Quelques mots de conclusion

Le taux de chômage est un indicateur **complexe à interpréter** :

- ▶ un indicateur parmi d'autres au sein de la constellation des indicateurs de conjoncture du marché du travail ;
- ▶ un indicateur qui dépend fortement de son processus d'élaboration, comme en témoigne l'impact du changement de questionnaire en 2013 ;
- ▶ un indicateur entâché d'une imprécision liée au sondage et à l'importance de non-réponse des ménages.

En pratique, **pour interpréter le taux de chômage (surtout en évolution) il convient de** :

- ▶ déterminer si l'évolution est significative au sens du sondage ou pas ;
- ▶ être au fait des changements de questionnaire récents susceptibles d'affecter le niveau de la série (et d'éventuelles rétopolations) ;
- ▶ savoir positionner cet indicateur par rapport à d'autres, pour comprendre la nature des transferts sur le marché du travail.

Le solde migratoire et le recensement de la population

Le solde migratoire et le recensement de la population

Point de départ : Bilan démographique 2019

Figure 1 - Évolution générale de la situation démographique

en milliers

	Population au 1 ^{er} janvier	Nombre de naissances vivantes	Nombre de décès	Solde naturel	Solde migratoire évalué	Ajustement ¹	Évolution de la population ² (en %)
2009	64 305	824,6	548,5	+ 276,1	+ 32		0,48
2014 hors Mayotte	65 907	811,4	558,7	+ 252,7	+ 30		0,43
2014 y.c. Mayotte	66 131	818,6	559,3	+ 259,3	+ 32		0,44
2015	66 422	798,9	593,7	+ 205,3	+ 40	- 65	0,37
2016	66 603	783,6	593,9	+ 189,8	+ 65	- 83	0,38
2017	66 774	769,6	606,3	+ 163,3	+ 46 _p	- 100 _p	0,31 _p
2018	66 884 _p	758,6	609,7	+ 148,9	+ 46 _p	- 101 _p	0,29 _p
2019	66 978 _p	753,0 _p	612,0 _p	+ 141,0 _p	+ 46 _p	- 101 _p	0,28 _p
2020	67 064 _p

p : donnée provisoire à la fin 2019 ; ... Résultat non disponible.

1. Du fait d'un changement de questionnaire du recensement de la population visant à améliorer la connaissance des situations de multi-résidence, un ajustement a été introduit pour estimer les évolutions de population à questionnement inchangé. Cet effet de questionnaire monte progressivement en charge sur plusieurs années [Insee, **Note technique**, 2020].

Le solde migratoire et le recensement de la population

Point de départ : Bilan démographique 2019

« Insee Première » : Format de publication le plus courant de l'Insee, 4 pages à **destination du grand public**. On a ici affaire à un « **marronnier** » annuel, dont la structure change peu d'une année sur l'autre.

Plusieurs thématiques Evolution de la population, fécondité, vieillissement, mariages et PACS. On ne s'intéresse ici qu'à l'**évolution de la situation démographique** → **équation démographique**.

Plusieurs données du document sont indiquées comme **provisoires** :

- ▶ la population au 1er janvier n'est **connue définitivement que jusqu'au 1er janvier 2017** (du fait du recensement de la population), les années suivantes sont estimées et seront révisées ;
- ▶ les naissances et les décès de la **toute fin d'année 2019** ne sont pas intégrés dans la publication faute de temps ;
- ▶ le solde migratoire est provisoire en 2017, 2018 et 2019 et **égal par convention ces trois années**.

Solde migratoire et équation démographique (1)

La formalisation sous-jacente à ce tableau est l'**équation démographique**.

Chaque année, l'évolution de la population se décompose en naissance, décès et solde migratoire :

$$\text{Evolution population} = \text{Naissances} - \text{Décès} + \text{Solde migratoire}$$

Dans cette équation, l'évolution de la population est donnée par le **recensement de la population**, le nombre des naissances et des décès par les **statistiques d'État-civil**.

Le **solde migratoire** est lui inconnu et en fait **déterminé par différence** :

$$\text{Solde migratoire} = \text{Evolution population} - \text{Naissances} + \text{Décès}$$

L'**estimation directe du solde migratoire est en effet complexe** : difficulté à enregistrer toutes les arrivées et surtout tous les départs (cas particuliers : îles).

Solde migratoire et équation démographique (2)

L'estimation par différence du solde migratoire est **risquée**, car celui-ci est **petit en niveau devant les estimations de population** et les composantes du solde naturel.

→ En particulier, une **rupture de faible ampleur** dans les estimations de population peut avoir un **impact très important sur le solde migratoire**.

C'est précisément ce qu'il s'est passé à partir du bilan démographique 2018 (recensement de la population 2016), en raison d'un **changement de questionnaire**.

Un ajustement a dû être intégré dans l'équation démographique pour que le solde migratoire fasse sens.

$$\text{Solde migratoire} = \text{Evolution population} - \text{Naissances} + \text{Décès} - \text{Ajustement}$$

Le solde migratoire et le recensement de la population

La source : le recensement de la population

Le recensement de la population est l'opération statistique **la plus importante menée par l'Insee** :

- ▶ 5 M de logements et 9 M d'individus recensés chaque année ;
- ▶ 8 000 communes concernées par an, 24 000 agents recenseurs sur le terrain.

La finalité du recensement de la population est double :

- ▶ avant toute chose, il permet de déterminer la **population légale de toutes les communes de France** au 1er janvier de chaque année : dotations, textes réglementaires, etc. ;
- ▶ par ailleurs, il constitue une des seules **sources d'information statistique à un niveau très fin** (communal voire infra-communal).

La méthode de recensement mise en œuvre en France depuis 2004 est **unique dans le monde** : ni recensement général (c'était le cas jusqu'en 1999), ni recensement par registre, mais **recensement rotatif et par sondage**.

Le principe du recensement rotatif et par sondage

Organiser périodiquement un **recensement général de la population** (par exemple tous les dix ans) est une opération extrêmement **lourde, coûteuse et surtout difficile à programmer** dans un contexte de réduction des dépenses publiques.

Inversement, le **coût d'un recensement à partir d'un registre est moindre et davantage lissé**, mais la France ne dispose pas de registre de population.

La solution conçue par l'Insee à la fin des années 1990 pour **lisser et alléger le coût du recensement** est la suivante :

- ▶ plutôt que d'interroger toutes les communes en une seule fois, celle-ci sont **réparties en cinq « groupes de rotation »** interrogés par roulement chaque année ;
- ▶ plus encore, dans les communes suffisamment grandes, des **techniques de sondage** peuvent être mobilisées pour limiter drastiquement le coût du recensement (interrogation d'un échantillon) **en introduisant une imprécision minimale.**

Le solde migratoire et le recensement de la population

Le recensement de la population dans les « petites » communes

Dans les communes de moins de 10 000 habitants (« petites » communes), le recensement est **exhaustif tous les cinq ans**.

L'Insee **doit cependant produire une population légale chaque année**, pas seulement celle correspondant au dernier recensement de la commune.

C'est cette nécessité qui induit un **décalage de presque trois ans** entre la date de parution et la date de validité des populations légales (populations légales au 1er janvier 2017 parues fin 2019).

Exemple Petite commune recensée en 2017

- ▶ au 1er janvier 2017 (fin 2019) : utilisation des données de collecte ;
- ▶ aux 1ers janvier 2018 et 2019 (fin 2020 et 2021) : mise à jour des données de collecte en suivant l'évolution du nombre de logements connus à la taxe d'habitation ;
- ▶ aux 1ers janvier 2020 et 2021 (fin 2022 et 2023) : interpolation entre le 1er janvier 2019 et la collecte 2022 (connue).



Le solde migratoire et le recensement de la population

Le recensement de la population dans les « grandes » communes

Dans les communes de 10 000 habitants ou plus (« grandes » communes), **environ 8 % des logements sont recensés chaque année.**

Plus précisément :

- ▶ la commune doit tenir à jour un **Répertoire des immeubles localisés** (Ril) ;
- ▶ les adresses du Ril sont réparties en **cinq groupes de rotation** enquêtés chaque année par roulement ;
- ▶ chaque année, **40 % des adresses du groupe de rotation sont tirées au sort.**

L'estimation de la population au 1er janvier de l'année N est obtenue en multipliant :

- ▶ le **nombre de logements au Ril** au 1er janvier N ;
- ▶ par le **nombre moyen de personnes par logement** estimé à partir des enquêtes de recensement N-2 à N+2.

En pratique donc, en grande commune la population légale est calculée sur la base d'une **enquête portant sur 40 % des logements de la commune.**

Du fait de son caractère non-exhaustif, le recensement rotatif présente, comme les enquêtes par sondage classique, une **certaine imprécision**.

En petite commune, cette imprécision est liée aux hypothèses formulées les années ne correspondant pas à la collecte exhaustive :

- ▶ qualité de la source utilisée pour mettre à jour la collecte ;
- ▶ pertinence de l'interpolation avec les données de la collecte suivante.

En grande commune, cette imprécision est liée à l'**estimation imparfaite du nombre de personnes par logement** :

- ▶ incertitude liée au sondage ;
- ▶ utilisation de données collectées à ± 2 ans d'intervalle.

Le solde migratoire et le recensement de la population

Quel impact de l'imprécision du recensement sur le solde migratoire ?

Pour des raisons théoriques, toutes ces formes d'imprécision ne **peuvent pas donner lieu à une quantification sous la forme d'un intervalle de confiance.**

Si on ne retient que l'**incertitude liée au sondage**, celle-ci conduit à une estimation précise à $\pm 15\ 800$ sur le champ de la population des ménages de France métropolitaine.

Cette imprécision n'est **pas négligeable devant le niveau du solde migratoire** (quelques dizaines de milliers d'individus) ; il convient donc d'en **tenir compte dans l'interprétation.**

Le solde migratoire et le recensement de la population

Le questionnaire du recensement : des logements, des personnes et des liens

Le questionnaire du recensement de la population est **court** :

- ▶ 1 page pour décrire le logement ;
- ▶ 2 pages pour décrire les liens entre personnes du logement ;
- ▶ un recto-verso par habitant permanent du logement.

Son objectif premier est de permettre un **dénombrement le plus exact possible de la population**, en excluant notamment les habitants dits non-permanents :

- ▶ **enfants multi-résidents** résidant moins de la moitié du temps dans le logement (garde partagée minoritaire) ;
- ▶ **étudiants majeurs** logés ailleurs pour leurs études ;
- ▶ **autres cas** : personne logées en institution, logement occasionnel pour le travail, etc.

La frontière entre habitants permanents et non-permanents est cependant **fragile** : une **évolution récente du questionnaire** s'est accompagnée d'une **rupture importante dans les estimations de population**.



La nouvelle analyse ménage-famille introduite en 2018 (1)

La nouvelle analyse ménage-famille (liens entre les occupants du logements) fait suite à plusieurs rapports du CNIS soulignant le **faible apport du recensement dans la mesure notamment des ruptures familiales**.

À cette occasion, la situation des enfants multi-résidents a été clarifiée et ils se sont trouvés bien davantage **mis en avant dans le questionnaire en tant qu'habitants non-permanents**.

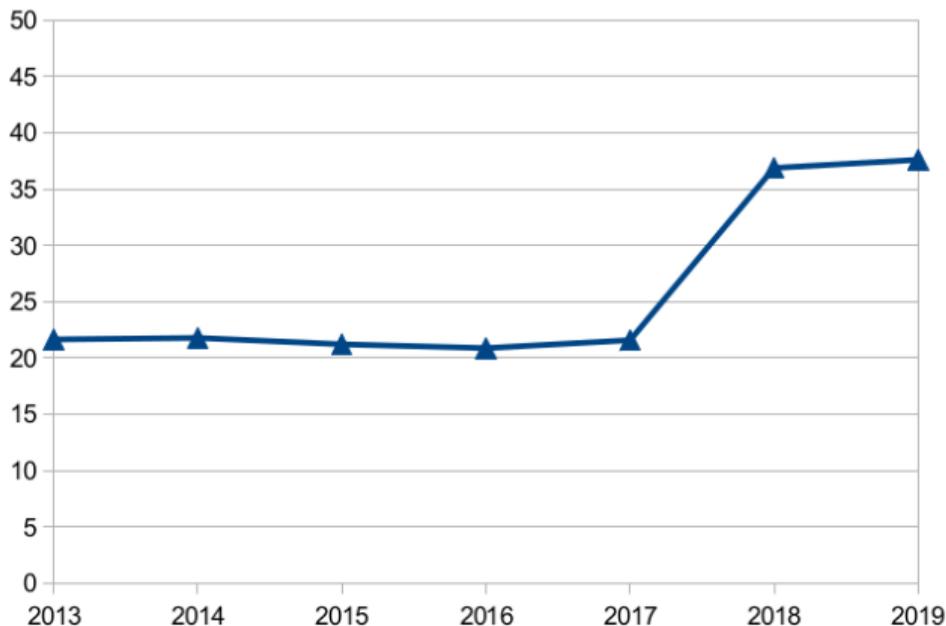
Cette évolution a induit une **augmentation très sensible du nombre d'enfants multi-résidents déclarés comme habitants non-permanents**, en parallèle d'une augmentation beaucoup plus faible qu'attendu de la population.

Hypothèse Des enfants multi-résidents auparavant déclarés (à tort) comme habitants permanents ne le sont désormais plus, d'où une **rupture dans l'évolution de la population**.



La nouvelle analyse ménage-famille introduite en 2018 (2)

Figure 2 : Nombre de personnes se déclarant en situation de multi-résidence sans être un habitant permanent du logement recensé (hors étudiants) – pour 1 000 résidences principales



Source : enquêtes annuelles de recensement de 2013 à 2019

Estimation d'un effet associé au changement de questionnaire à partir de la rupture de série apparente : **-501 000 individus en cycle RP complet, soit -0,75 % de la population.**

Etant donnée la méthodologie du recensement, cet effet n'est pas brutal mais se fera sentir sur les **résultats des RP 2016 à 2022 inclus** (fin 2024).

Étant donné le caractère légal des estimations de population, il a été décidé de **ne pas rétropoler les séries de population.**

En revanche l'effet questionnaire est **introduit sous la forme d'un ajustement dans le bilan démographique**, de sorte que les évolutions de population, et donc le solde migratoire, aient un sens.

Le solde migratoire et le recensement de la population

Les évolutions du protocole de collecte

La collecte du recensement de la population obéit à un protocole précisément défini et assez peu employé par ailleurs : le **dépôt-retrait**.

Le développement de la **collecte par internet** à partir de 2015 a singulièrement allégé la charge de collecte (pas de retrait) et de traitement en aval (pas de saisie optique).

Toute évolution est néanmoins susceptible d'introduire des ruptures de série :

- ▶ interrogations sur un **effet de mode éventuel sur l'emploi** ;
- ▶ réaffirmation du mode auto-administré par rapport à un mode en **face-à-face sur tablette** ;
- ▶ réflexions en cours sur la collecte dite séquentielle (dépôt d'une notice de connexion dans la boîte aux lettres), évaluation du risque d'un **impact sur le taux de résidences non-principales**.

Quelques mots de conclusion

Le solde migratoire est un **indicateur souvent très commenté**, notamment par des acteurs politiques.

Ses modalités d'élaboration doivent cependant conduire à une **certaine prudence dans les interprétations** :

- ▶ il est obtenu par différence avec des quantités d'ordres de grandeurs sensiblement supérieurs et **potentiellement imprécises** ;
- ▶ il est publié à partir de **projections et d'hypothèses** pas toujours bien connues ;
- ▶ il intègre actuellement un **ajustement** pour tenir compte d'une évolution importante du questionnaire du recensement de la population, ajustement dont l'évaluation est elle-même imprécise.