

Séminaire Fabrique des données

Comment se fabrique et s'utilise une enquête de la statistique publique ? L'exemple de l'enquête Emploi

L'ÉCOLE
DES HAUTES
ÉTUDES EN
SCIENCES
SOCIALES

PSL 
RESEARCH UNIVERSITY PARIS

Martin CHEVALIER (DREES)

Année universitaire 2022-2023

Objectifs de la séance

Mettre en avant les **spécificités** des enquêtes menées par la statistique publique en France.

Donner des **points de repère** et des éléments de **culture générale sur le système statistique public**, son organisation et ses méthodes.

Présenter plus en détails un dispositif majeur de la statistique publique, **l'enquête Emploi en continu (EEC)**.

Introduction

Plan de la présentation

Cadre institutionnel

Collecte

Méthodologie de sondage

Diffusion

L'EEC : un dispositif majeur de la statistique publique

Quelques mots sur mon parcours (ou encore : d'où je parle)

Sociologue de formation, statisticien Insee depuis 2014 :

- ▶ 2014-2018 : méthodologue à la **Division Sondages** de l'Insee, en charge du calcul de précision dans les enquêtes ménages ;
- ▶ 2018-2021 : responsable de la méthodologie du **Recensement de la population** à l'Insee ;
- ▶ depuis 2021 : adjoint du bureau Lutte contre l'exclusion de la **DREES** (service statistique des ministères de la santé et des solidarités).

Expertise en **méthodologie d'enquête** (sondage surtout) + vision des **pratiques en dehors de l'Insee** avec mon poste actuel à la DREES.

Cadre institutionnel

Qu'est-ce qu'une enquête de la statistique publique ?

Une enquête de la statistique du publique est une enquête (co)produite par une des **institutions du service statistique public** (SSP) français :

- ▶ Institut national de la statistique et des études économiques (Insee) ;
- ▶ 16 services statistiques ministériels (SSM) : Dares (Travail), DREES (Santé-Solidarités), Depp (éducation), SDeS (environnement), etc.

Ces enquêtes sont très souvent réalisées en collaboration avec d'**autres institutions hors SSP** : Institut national d'étude démographique (Ined), Santé publique France (SPF), CNRS, etc.

Au 1er janvier 2022, le SSP compte environ **6 500 agents**.

Le secret statistique

Les enquêtes de la statistique publique sont couvertes par le **secret statistique**, défini par la loi du 7 juin 1951.

En dérogation aux règles communes applicables au secret professionnel, le secret statistique est **opposable** à toute réquisition judiciaire ou émanant d'autorités administratives (fiscale ou douanière par exemple).

Il exclut que des données collectées à des fins statistiques puissent être utilisées ou réutilisées **à des fins de décision à l'égard des personnes concernées**. En particulier, il exclut strictement toute utilisation ou réutilisation de ces données à des fins de **contrôle fiscal ou de répression économique**.

Les données protégées par le secret statistique **peuvent néanmoins être communiquées** en réponse à des demandes formulées à des fins statistiques ou de recherche scientifique ou historique d'**intérêt public**.

Parenthèse : la protection des données personnelles

Les obligations en matière de protection des données personnelles ne sont **pas spécifiques** aux enquêtes de la statistique publique.

Deux textes font **référence en la matière** :

- ▶ la loi Informatique et libertés de 1978, qui crée en particulier la CNIL ;
- ▶ le Règlement général sur la protection des données (RGPD), qui est entré en vigueur en 2018.

Plusieurs **principes fondamentaux** : finalité (pour la statistique publique : mission d'intérêt public), proportionnalité et pertinence, durée de conservation limitée, sécurité et confidentialité, respect du droit des personnes.

Une **analyse d'impact en protection des données** (AIPD, PIA en anglais) est souvent nécessaire et détaille les mesures à mettre en œuvre.

Le rôle du CNIS dans la statistique publique

Le conseil national de l'information statistique (CNIS) joue un **rôle central au sein du SSP** :

- ▶ porter la **demande sociale** : composé principalement de représentants syndicaux et patronaux ;
- ▶ **coordonner** le SSP : programmation des travaux, avis d'opportunité sur les nouvelles enquêtes ;
- ▶ garantir la **qualité** et rendre la réponse à l'enquête obligatoire : comité du label ;
- ▶ accorder l'accès aux **données sensibles** : comité du secret statistique.

Le passage devant les différentes instances du CNIS a des **conséquences importantes** sur le déroulement concret d'une enquête.

Le contexte européen

Les instituts nationaux de statistique (INS) sont **coordonnés au niveau européen par Eurostat** :

- ▶ adoption d'un **référentiel commun de qualité** : code de bonnes pratiques de la statistique européenne, revues par les pairs ;
- ▶ constitution d'un **socle réglementaire** pour les enquêtes : **fondement juridique** et **harmonisation**.

Ce contexte a un **impact très concrets** sur les enquêtes :

- ▶ enquêtes **obligatoires** : garantie de réalisation pour les enquêtes concernées mais plus de difficultés pour mener d'autres enquêtes ;
- ▶ **contraintes** sur les questionnaires, les méthodes et les délais (*timeliness*) ;
- ▶ **reporting** et transmission de données *ad hoc* : rapport qualité, hypercubes, etc.

Financement et comitologie

Une grande enquête de la statistique publique représente un coût de **plusieurs millions d'euros**, souvent **partagé entre plusieurs institutions** (thèmes ou sous-échantillons spécifiques).

La **comitologie** de l'enquête (comités de pilotage et de suivi, conseil scientifique) vise à **représenter** les institutions parties prenantes mais aussi les chercheurs du domaine et les acteurs institutionnels concernés.

Les décisions structurantes pour l'enquête (notamment sur le questionnaire) résultent d'un **équilibre** entre :

- ▶ les **besoins** exprimés par les acteurs sollicités... : reconduction de questions déjà posées mais aussi nouveaux questionnements ;
- ▶ ...et les **contraintes** pesant sur la réalisation de l'enquête : durée du questionnaire, taille de l'échantillon, disponibilité du réseau enquêteur.

Collecte

Des moyens importants pour de « grandes enquêtes »

Les moyens engagés dans les enquêtes de la statistique publique sont **importants**, ce qui a plusieurs conséquences :

1. Les échantillons peuvent être de **grande taille** : 10 000 à 20 000 personnes pour une enquête « classique », 110 000 chaque trimestre pour l'enquête Emploi en continu, plus de 350 000 pour l'enquête Epicov.
2. Cette taille d'échantillon importante permet en particulier de définir des **sous-échantillons spécifiques** (surreprésentations) avec potentiellement des questionnements spécifiques ;
3. La **collecte en face-à-face** est possible, ce qui permet des questionnements particulièrement longs ou complexes (enquête Histoire de vie et patrimoine).

Collecte

La possibilité de mobiliser le réseau d'enquêteurs de l'Insee

Les enquêtes de la statistique publique peuvent **mobiliser le réseau des 800 enquêtrices et enquêteurs de l'Insee** répartis sur tout le territoire :

La stabilité du réseau et la formation importante des enquêtrices et enquêteurs avant chaque enquête permet de **collecter des enquêtes complexes** :

- ▶ **thématique complexe** : enquête Histoire de vie et patrimoine, enquête Cadre de vie et sécurité ;
- ▶ **protocole complexe** : enquête Sans-domicile ;
- ▶ **délais de collecte très courts** : enquête Emploi en continu (*cf. infra*).

Remarque Les SSM peuvent aussi recourir au réseau d'enquêteurs de grands instituts de sondage (Ipsos notamment).



Des contraintes néanmoins importantes

Diminution structurelle des moyens dans la sphère publique :

focalisation sur les enquêtes/questions adossées à des règlements européens, recours accru à la collecte internet (au prix d'éventuels **effets de mode**).

Gestion très contrainte des moyens enquêteurs : moyens limités, programmation fine des enquêtes plusieurs années à l'avance.

Outils spécifiques pour le codage du questionnaire : filière informatique spécifique pour prendre en charge le multimode en cours d'élaboration.

Méthodologie de sondage

La possibilité de mener un échantillonnage probabiliste

La statistique publique est un des rares cas d'application des **techniques d'échantillonnage probabiliste**.

Ces techniques requièrent en effet de disposer de **bases de données** (raisonnablement) exhaustives susceptibles d'être utilisées pour **constituer un échantillon de personnes tiré au sort**.

La constitution et l'exploitation de ces bases de données est une des **missions historique de l'Insee** : recensement de la population, exploitation de fichiers administratifs (données fiscales notamment).

Cette particularité **affecte de nombreux aspects** de la méthodologie d'une enquête.



Exploitation de l'information auxiliaire disponible dans les fichiers

Au-delà de l'information de contact, les bases de données utilisées par l'Insee pour tirer les échantillons des enquêtes contiennent en général des **informations corrélées aux thèmes de l'enquête**.

Exemples

- ▶ localisation géographique pour l'enquête nationale sur le logement 2013 ;
- ▶ asujettissement à l'ISF pour l'enquête Patrimoine 2014.

Au-delà des méthodes classiques de **stratification** (*cf.* dernières séances du cours ITS au S1), des méthodes plus sophistiquées d'**échantillonnage équilibré** ont été développées à l'Insee pour tirer parti de cette information.



Prise en compte du réseau enquêteur : l'échantillon-maître

Dans le cadre d'interrogations en face-à-face, il est essentiel de **limiter au maximum les déplacements des enquêteurs**, qui représentent une part importante du coût de l'enquête.

Un ensemble de méthodes de sondage ont été développées pour répondre à ce type bien spécifique de contraintes, l'échantillonnage à **plusieurs degrés** : on tire **d'abord des zones** (par exemple des communes) puis des logements au sein de ces zones.

En pratique à l'Insee, ce principe est mis en œuvre *via* l'**échantillon-maître** : les zones dans lesquelles les individus seront enquêtés sont **tirées au sort une fois pour toute tous les dix ans**.

Traitements post-collecte sophistiqués

Le cadre probabiliste permet également la mise en œuvre de **traitements post-collecte sophistiqués**, dont les propriétés statistiques sont **bien établies** :

- ▶ **correction de la non-réponse** par repondération afin de limiter l'ampleur du **biais de non-réponse** ;
- ▶ **calage sur marges** pour **améliorer la précision des estimateurs** calculés à partir de l'enquête.

L'ensemble des opérations d'échantillonnage et de traitement post-collecte peuvent être prises en compte dans le cadre d'un **calcul de précision**, notamment pour répondre aux **demandes d'Eurostat**.

Des méthodes d'estimations « **petits domaines** » peuvent également être utilisées dans des cas particuliers (carroyage du recensement par exemple).

Diffusion

Exploitation primaire par le SSP

Les enquêtes de la statistique publique font toujours l'objet d'une **exploitation primaire** par leur(s) producteur(s), sous des formats variés.

Chaque organisme du SSP dispose ainsi de **formats de publications adaptés à la nature des informations à transmettre** :

- ▶ publications **rapides** : Informations rapides (Insee), Dares indicateurs ;
- ▶ publications **courtes** : Insee première, études et résultats (DREES) ;
- ▶ jeux de données statistiques : data.drees ;
- ▶ publications **plus longues au format en général moins contraint** : dossiers de la DREES, documents d'études (Dares) ;
- ▶ **ouvrages de référence** : Insee références, panoramas de la DREES.

Ces publications sont toujours **gratuites et librement accessibles sur internet**.



Diffusion

Exploitation secondaire par des chercheurs ou des étudiants

Les données des enquêtes de la statistique publique sont **mises à disposition des chercheurs, des étudiants voire du grand public** par plusieurs biais :

- ▶ données ouvertes en **open data** : data.gouv.fr, fichier détail du recensement de la population sur le site de l'Insee ;
- ▶ fichiers de production et de recherche (FPR) anonymisés accessibles gratuitement aux chercheurs et étudiants *via* le **réseau Quételet** ;
- ▶ fichiers contenant des **informations sensibles** (opinions politiques, etc.) ou **fortement réidentifiantes** (localisation fine, identifiants en vue d'appariements) accessibles *via* le **Centre d'accès sécurisé distant** (CASD), après avis du comité du secret et moyennant paiement d'un abonnement (onéreux).

Les SSM **financent régulièrement** des doctorats, post-doctorats ou projets de recherche pour exploiter les données de leurs enquêtes.

Diffusion

Appariements de données d'enquête avec des données administratives

Les organismes du SSP ont la possibilité d'**appairer ou de faire appairer** par l'Insee les données d'enquêtes avec d'**autres données disponibles par ailleurs**.

Exemple Enquête auprès des Bénéficiaires de minima sociaux (BMS) 2018 : appariement par l'Insee pour le compte de la DREES pour récupérer des **informations socio-ficales** (impôts, CAF, etc.) et ainsi pouvoir calculer le **taux de pauvreté monétaire** des personnes bénéficiaires de minima sociaux (RSA, AAH, etc.).

La généralisation du **code statistique non-signifiant** (CSNS) au sein du SSP vise à **simplifier et à sécuriser** ces travaux d'appariements.

L'EEC : un dispositif majeur de la statistique publique



Cadre institutionnel

L'enquête Emploi en continu est une **pierre angulaire** de la statistique publique.

Elle est la déclinaison française des **Labour force surveys européennes** (LFS) et est encadrée par le règlement IESS (*Integrated european social statistics*).

Son principal objectif est de mesurer le **taux de chômage** au sens du Bureau international du travail (BIT), mais son questionnaire riche permet d'aborder de **nombreuses thématiques** : temps de travail, conditions de travail, revenus, mobilité sociale.

En conséquence, son questionnaire est **long et complexe** et organisé en modules (module A : position sur le marché du travail au sens du BIT).

Refonte du questionnaire de l'enquête Emploi en continu en 2013.

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](#)

L'EEC : un dispositif majeur de la statistique publique

Parenthèse : De l'importance de la formulation 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 ?

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

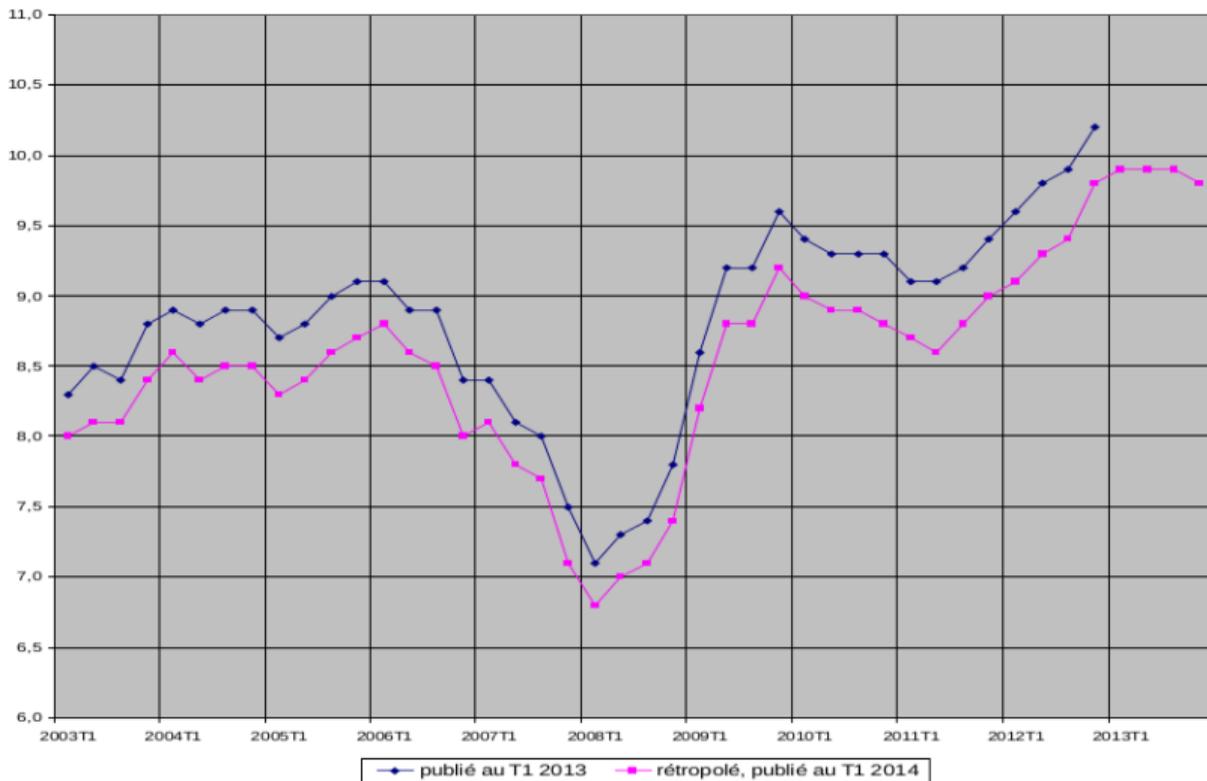
- ▶ (-) reformulation de la question sur la recherche d'emploi, moins de confusion avec le souhait de travailler ;
- ▶ (-) suppression de la modalité "Pôle emploi" ;
- ▶ (-) ajout de la question sur le travail informel ;
- ▶ (+) nouvelles modalités de recherche d'emploi (internet) ;
- ▶ (+) reformulation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

Conséquences pratiques

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

Parenthèse : De l'importance de la formulation des questions (4)

Taux de chômage cvs, France métropolitaine, en %



Collecte

Plusieurs éléments **singularisent la collecte de l'EEC** :

- ▶ **gros échantillon** : 80 000 logements par trimestre, soit cinq fois plus qu'une enquête aperiodique classique ;
- ▶ **en continu** depuis 2003 : des logements sont interrogés chaque semaine ;
- ▶ **délais de collecte très courts** : interrogation au plus deux semaines et deux jours après la semaine de référence ;
- ▶ **panel rotatif** : les logements sont interrogés six trimestres de suite puis sortent de l'échantillon (et sont remplacés par d'autres) ;
- ▶ **multimode de longue date** : première interrogation en face-à-face, les suivantes par téléphone (et désormais internet).

L'EEC est un **dispositif majeur pour le réseau d'enquêteurs de l'Insee**, dont elle représente une part substantielle de l'activité.

L'EEC : un dispositif majeur de la statistique publique

Méthodologie (1) : tirage aréolaire et coordonné

Le plan de sondage l'EEC est adapté aux **contraintes de collecte fortes de l'opération** :

- ▶ **tirage aréolaire** : afin de rendre possible une collecte dans des délais très courts, les logements tirés sont littéralement voisins (on tire non des logements mais des « aires » géographiques à enquêter en totalité) ;
- ▶ **tirage coordonné avec l'échantillon-maître** : afin de limiter les distances à parcourir pour les enquêteurs, le tirage garantit que les aires à enquêter pour l'EEC sont **proches des communes de l'échantillon-maître**.

Méthodologies de constitution des aires de collecte et algorithmes de tirage **particulièrement complexes**, afin de **garantir la meilleure précision possible aux estimateurs construits** tout en respectant les contraintes de collecte.

L'EEC : un dispositif majeur de la statistique publique

Méthodologie (2) : traitements post-collecte

Traitements post-collecte **relativement standards et automatisés** :

- ▶ **correction de la non-réponse par repondération** ;
- ▶ **calage sur des marges de population** calculées à partir du recensement de la population (trimestrialisé).

Investissement dans le **calcul de précision** :

- ▶ **forte demande** au niveau européen : rapport qualité, bornes fixées par le règlement IESS ;
- ▶ méthodologies de tirage et d'estimation à prendre en compte **le plus exhaustivement possible** ;
- ▶ **complexité des indicateurs** : évolution d'un trimestre à l'autre, moyennes annuelles, etc.

L'EEC : un dispositif majeur de la statistique publique

Parenthèse : Des pondérations fréquemment mises à jour

En raison des particularités du recensement de la population français, les estimations de population servant de marges de calage ne sont **définitives que trois ans après l'année de référence**.

Un **nouveau calage sur marges** est alors effectué sur ces marges et les données de l'EEC sont **rediffusées avec un poids mis à jour**.

Il est d'autant plus important de procéder à cette mise à jour que **l'EEC sert elle-même à calculer les marges de calage** de très nombreuses enquêtes de la statistique publique.

Remarque La **forte réévaluation** de la population française suite à l'évolution du questionnaire du recensement en 2018 a ainsi induit une rétopolation des estimations de l'EEC jusqu'à 2014 inclus.

L'EEC : un dispositif majeur de la statistique publique

Diffusion (1) : le taux de chômage

Publication du **taux de chômage** environ **six semaines** après la fin du trimestre (indicateur sur calendrier).

**Au quatrième trimestre 2022,
le taux de chômage est quasi
stable à 7,2 %**

Chômage au sens du BIT et indicateurs sur le
marché du travail (résultats de l'enquête
emploi) (BIT) - quatrième trimestre 2022

INFORMATIONS RAPIDES

N° 37

Paru le : 14/02/2023

Prochaine parution : 17/05/2023 à 07h30 - premier
trimestre 2023

> [Découvrir la collection](#)

VERSION IMPRIMABLE

(pdf, 1 Mo)



Autres utilisations par l'Insee :

- ▶ **estimations d'emploi** au niveau local (jusqu'à la zone d'emploi) ;
- ▶ **Insee références annuel** sur le marché du travail ;
- ▶ études ponctuelles.



Historiquement, les fichiers de données de l'EEC étaient **diffusés directement sur le site de l'Insee** : ce n'est aujourd'hui plus le cas.

Comme la plupart des enquêtes de l'Insee, les données de l'EEC sont diffusées selon deux canaux :

- ▶ le réseau Quételet pour le **fichier de production et de recherche (FPR)** ;
- ▶ le CASD pour la base dite « Z », *i.e.* la **dernière base de production en interne Insee** (avec des informations de localisation plus fines).

L'EEC : un dispositif majeur de la statistique publique

Diffusion (3) : ERFS, l'EEC appariée avec les données fiscales

Une modalité de diffusion extrêmement structurante des données de l'EEC est l'**enquête Revenus fiscaux et sociaux** (ERFS).

Il s'agit des **données du quatrième trimestre de l'EEC d'une année donnée** appariées avec les **données des impôts et des différentes caisses de sécurité sociale**.

Ce sont ces données qui permettent de calculer le **taux de pauvreté monétaire** de la population ainsi que des analyses fines de **l'impact du système socio-fiscal en termes de redistribution**.

Tableau 1 Effet de chaque étape de la redistribution sur le taux de pauvreté, l'intensité de la pauvreté et le seuil de pauvreté à 60 % du niveau de vie médian, en 2019

	Taux de pauvreté		Intensité de la pauvreté		Seuil de pauvreté mensuel	
	Niveau (en %)	Effet en niveau (en points)	Niveau (en %)	Effet en niveau (en points)	Niveau (en euros)	Effet en niveau (en euros)
Revenu initial	22,2	-	38,5	-	1 190	-
Impôts directs ¹	21,3	-0,9	38,7	+0,2	1 057	-133
Prestations familiales ²	19,4	-1,9	32,9	-5,8	1 086	+29
Allocations logement	17,7	-1,7	27,9	-5,0	1 089	+3
Minima sociaux ³	16,0	-1,7	21,3	-6,6	1 095	+6
Garantie jeunes	15,9	-0,1	21,0	-0,3	1 095	0
Prime d'activité	14,6	-1,3	19,7	-1,3	1 102	+7
Revenu disponible	14,6	-7,6	19,7	-18,8	1 102	-88

Comment se fabrique et s'utilise une enquête de la statistique publique ? L'exemple de l'enquête Emploi

De nombreux éléments **singularisent les enquêtes de la statistique publique** dans le paysage des enquêtes par sondage :

- ▶ **cadre institutionnel** : secret statistique, contexte européen, etc. ;
- ▶ **collecte** : gros échantillons, réseau d'enquêteurs spécialisés ;
- ▶ **méthodologie** : échantillonnage probabiliste, méthodologie sophistiquée ;
- ▶ **diffusion** : principe de gratuité, pour les études et les données.

L'enquête Emploi en continu (EEC) est un **dispositif majeur du SSP** :

- ▶ cadre réglementaire extrêmement **strict** ;
- ▶ **forte empreinte** sur le réseau enquêteur, qu'elle contribue à structurer ;
- ▶ **investissement méthodologique** important ;
- ▶ **deux indicateurs-phares** : le taux de chômage et – *via* ERFS – le taux de pauvreté monétaire.

Comment se fabrique et s'utilise une enquête de la statistique publique ?

L'exemple de l'enquête Emploi

Merci de votre attention !

Martin Chevalier (DREES)
martin.chevalier@sante.gouv.fr