
L'enquête Emploi en Continu 2019-2028 : étude du second degré du plan de sondage

Thomas SAUVAGET (*), Thomas DEROYON (**), Thomas MERLY-ALPA (***)

(*) SSP75, Sous-direction des synthèses statistiques et des revenus

(**) Drees, Sous-direction de l'observation de la santé et de l'assurance maladie

(***) Ined, Service des enquêtes et des sondages

thomas.sauvaget1@agriculture.gouv.fr, thomas.deroyon@sante.gouv.fr,
thomas.merly-alpa@ined.fr

Mots-clés : Échantillonnage, équilibrage, variance, collecte, Enquête Emploi en Continu.

Domaines : 1.2 échantillonnage particulier.

Résumé

L'Enquête Emploi en Continu (EEC) est la seule source fournissant une mesure des concepts d'activité, de chômage, d'emploi et d'inactivité tels qu'ils sont définis par le Bureau International du Travail (BIT). Elle est réalisée en continu, chaque semaine, depuis 2003. Chaque trimestre, l'EEC interroge les occupants d'un échantillon d'environ 70 000 logements (ce qui représente environ 100 000 répondants) sur leur situation sur le marché du travail. Cet échantillon vise à décrire l'ensemble des personnes de 15 ans et plus vivant en logement ordinaire (i.e. hors communautés comme les maisons de retraite, les prisons, les casernes...) à titre de résidence principale.

L'échantillon de l'EEC est un échantillon de logements, dont tous les occupants dans le champ de l'enquête sont interrogés chaque trimestre, en face à face puis par téléphone et internet. Cet échantillon est aréolaire : les logements à enquêter sont regroupés en « grappes » d'unités contigües (en tout cas proches géographiquement). Il est également rotatif : les grappes sont regroupées en secteurs, regroupant six grappes proches géographiquement. Les logements d'une grappe sont interrogés pendant six trimestre consécutifs, puis sont remplacés par les logements d'une autre grappe d'un même secteur. L'échantillon de l'EEC est donc utilisé pendant trente-six trimestres, et renouvelé tous les neuf ans.

Sa constitution suit schématiquement deux étapes :

- une première étape de constitution de zones géographiques restreintes, appelées « secteurs », et de tirage des 2 944 secteurs où se situera l'enquête ;
- une seconde étape de tirage des grappes de logements à enquêter dans chaque secteur.

Entre 2017 et 2019, un travail de refonte du plan de sondage a été mené à l’Insee par la Division Sondages. Dans un premier temps, le choix de ces secteurs, premier degré du plan de sondage, a été réalisé en 2018 de manière coordonnée avec le nouvel Échantillon-Maître de l’Insee, dans le cadre du projet Nautile, par tirage spatialement équilibré.

La difficulté liée à la seconde étape est issue du mode de construction des secteurs : en effet, les secteurs n’avaient pas été construits en mobilisant l’ensemble des logements, mais seulement les Résidences Principales (RP) présentes en métropole dans le prototype Fidéli 2016 et exploitables, soit 28 282 687 RP. L’analyse des résultats de collecte de l’EEC sur les périodes récentes montra cependant que le taux de conversion des Résidences Non Principales (RNP : résidences secondaires et logements vacants) en RP, i.e. la part de RNP de la base de sondage devenues des RP au moment de la collecte, était loin d’être négligeable. Aussi, il a été décidé de continuer à échantillonner des RNP dans l’EEC comme dans le plan de sondage précédent.

Dans cet article, on décrit ainsi pourquoi et comment, en 2019, les étapes de construction du second degré du plan, à savoir l’obtention des grappes de logements à enquêter et leur pondération initiale, ont nécessité :

- le rattachement du stock de RNP 2016 aux différentes grappes de RP, constituant ainsi ce que l’on appelle des « grappes-univers » (du millésime 2016), chacune étant un univers de sondage dans lequel sera échantillonnée la grappe définitive ;
- la mise à jour des statuts des logements (RP ou RNP) entre 2016 et 2018 ;
- le rattachement des logements nouveaux entre 2016 et 2018 et la suppression des logements détruits ;
- la sélection dans les grappes-univers 2018 ainsi obtenues, de grappes à collecter satisfaisant des contraintes sur leurs nombres de RP et de RNP.

Une des contraintes importantes pour la collecte étant d’enquêter tous les logements d’un même étage, la sélection mentionnée a consisté en un tirage d’étages plutôt que directement de logements.

Les études méthodologiques menées ont abouti à choisir un tirage stratifié à deux phases sous contraintes de variance additionnelle minimale de la seconde phase et allocations strictement positives. La pondération obtenue reflète les différentes étapes du tirage. Le nouvel échantillon de l’EEC a été progressivement déployée dans les nouveaux secteurs géographiques à partir du T3 2019.

Abstract

In France, the renewed Labour Force Survey (LFS) has progressively been deployed across new geographical areas, starting in the third trimester of 2019. The choice of those areas, which constitutes the first stage of the sampling frame, has been realised in 2018, in a coordinated fashion with the new Master Sample of Insee (within the Nautile project). In 2019, the design of the second stage of the sampling frame, namely the construction of the clusters of households to be surveyed, as well as the overall weighting, has been realized. The methodology chosen for the construction of that second stage is the topic of this communication.

Introduction

L’EEC est la seule source fournissant une mesure des concepts d’activité, de chômage, d’emploi et d’inactivité tels qu’ils sont définis par le Bureau international du travail (BIT). Elle est

réalisée en continu, chaque semaine, depuis 2003. Chaque trimestre, l’EEC interroge les occupants d’un échantillon d’environ 70 000 logements (ce qui représente environ 100 000 répondants) sur leur situation sur le marché du travail. Cet échantillon vise à décrire l’ensemble des personnes de 15 ans et plus vivant en logement ordinaire (i.e. hors communautés comme les maisons de retraite, les prisons, les casernes...) à titre de résidence principale.

L’échantillon de l’EEC est un échantillon de logements, dont tous les occupants dans le champ de l’enquête sont chaque trimestre interrogés. Cet échantillon est aréolaire : les logements à enquêter sont regroupés en « grappes » d’unités contiguës (en tout cas proches géographiquement). Il est également rotatif : les grappes sont regroupées en secteurs, regroupant six grappes proches géographiquement. Les logements d’une grappe sont interrogés pendant six trimestre consécutifs, puis sont remplacés par les logements d’une autre grappe d’un même secteur. L’échantillon de l’EEC est donc utilisé pendant 36 trimestres, et renouvelé tous les neuf ans. Sa constitution suit schématiquement deux étapes :

- une première étape de choix de zones géographiques restreintes, appelées « secteurs », où se situera l’enquête ;
- une seconde étape de construction des ensembles de logements à enquêter.

Pour la construction des secteurs utilisés sur la période 2009-2018, voir Loonis [3]. L’échantillon de l’EEC arrivant à échéance en 2018, la question de son renouvellement s’est posée (voir Guillaumat-Tailliet et Tavan [2] et Quénechdu et Tchobanian [5] pour le contexte plus large de la rénovation de l’enquête emploi). Les nouveaux secteurs tirés en 2018 pour la période 2019-2028 sont spatialement proches des Unités Primaires (UP) du nouvel Échantillon-Maître (EM) de l’Insee. Ceci afin de créer de grandes zones d’activité pour les enquêteurs, et ainsi de simplifier l’organisation du travail. Du fait du changement de questionnaire de l’EEC en 2021, deux échantillons de secteurs ont été tirés : 2 944 pour l’échantillon principal, et 734 pour un échantillon supplémentaire qui permettra d’appréhender les ruptures de série. Nous renvoyons à Costa *et al.* [1] et à Merly-Alpa *et al.* [5] au sujet de la coordination entre EEC et EM, ainsi que sur les constructions mentionnées dans la suite de cette section.

Les logements dans le champ de l’enquête sont les logements ordinaires (i.e. hors communautés) résidences principales au moment de la collecte. Il existe cependant un décalage systématique entre la base de sondage et la population au moment de la collecte : des logements occupés à titre de résidence principale dans la base de sondage sont vacants au moment de la collecte, des logements vacants ou des résidences secondaires au moment de la constitution de la base de sondage sont devenus des résidences principales au moment où l’enquête est collectée. Naturellement, la part de logements qui sont effectivement des résidences principales au moment de la collecte est largement supérieure parmi les logements qui sont également des résidences principales dans la base de sondage, et ce d’autant plus que la collecte a lieu peu après le tirage de l’échantillon. Inclure des résidences non principales dans l’échantillon conduit donc à interroger des logements dont la probabilité d’être dans le champ est relativement faible et peut conduire à consacrer des moyens de collecte rares à des logements finalement hors-champ.

Aussi, dans la plupart des enquêtes auprès des ménages, la base de sondage est limitée aux habitations occupées à titre de résidence principale. L’échantillon de l’enquête Emploi inclut par contre historiquement des résidences non principales et des logements vacants dans la base de sondage. L’échantillon de logements interrogés au cours des six trimestres dans une grappe est en effet déterminé au moment où cette grappe est mise en collecte. Au fur et à mesure, la distance entre la base de sondage et la réalité du terrain s’accroît, et la part de RNP devenues RP risque d’augmenter, accroissant de fait le défaut de couverture pour l’EEC. Ceci garantit également que les estimateurs calculés à partir de l’enquête Emploi ne souffrent pas du défaut de couverture lié à la non inclusion des RNP : ils peuvent donc être utilisés dans un calage sur marges pour corriger ce défaut de couverture dans les autres enquêtes.

Au moment où les réflexions sur la constitution du nouvel échantillon de l'EEC ont débuté, il n'était pas encore certain que l'intégration de RNP dans l'échantillon serait maintenue. Aussi les secteurs et les grappes de l'EEC ont d'abord été constitués à partir des seules résidences principales de Fidéli 2016. L'analyse des résultats de collecte de l'EEC sur les périodes récentes a cependant montré que le taux de conversion des RNP en RP (la part de RNP de la base de sondage devenues des RP au moment de la collecte) était loin d'être négligeable (voir supra), aussi il a été décidé de continuer à échantillonner des RNP dans l'EEC. Les RP demeurent cependant prioritaires : les taux de sondage appliqués aux RP doivent être plus faibles que ceux appliqués aux RNP.

Dans un premier temps, les Résidences Principales (RP) présentes en métropole dans le prototype Fidéli 2016 et exploitables, soit 28 282 687 RP, ont été réparties en « étages », au sens statistique : il s'agit soit de logements situés au même étage d'un bâtiment, soit de maisons indépendantes.

Une agrégation des étages sous contraintes de taille a permis d'aboutir à des grappes de RP : elles comptent 20,2 RP en moyenne, dont 80,15 % entre 19 et 21 RP, et la quasi-totalité entre 17 et 24 RP.

Puis, les grappes de RP ont été regroupées par 6 pour former les secteurs¹. Cependant, il restait encore plusieurs étapes avant de disposer des grappes de logements effectivement mises en collecte à partir du 3e trimestre 2019, comme le présente la section suivante.

1 Construction de grappes-univers

Il est tout d'abord nécessaire de rattacher aussi toutes les Résidences Non Principales (RNP : résidences secondaires et logements vacants) aux grappes de RP. Ceci provient du constat qu'en collecte terrain de l'EEC actuelle, de nombreuses RNP s'avèrent être des RP. Écarter les RNP entraînerait donc un défaut de couverture. En 2017-18, le taux de conversion moyen de l'état « RNP » à « résidence principale d'après la collecte » était de 34 % (il monte jusqu'à 61 % en Ile-de-France dans les cas de grappes ayant moins de 6 RNP Fidéli).

Par ailleurs, à chaque changement de millésime Fidéli N à N+1 on observe que :

- des logements disparaissent (nous les appellerons « logements disparus ») : ils regroupent à la fois des logements physiquement détruits et des logements qui quittent temporairement la base de sondage (ces derniers sont soit hors du triplet Principal / Secondaire / Vacant, soit des logements temporairement non-traitables, par exemple sans coordonnées XY) ;
- des logements apparaissent (nous les appellerons « nouveaux logements ») : ils regroupent à la fois des nouvelles constructions, mais aussi des réaménagements (maison découpée en appartements), et des réapparitions (logement connu dans un millésime antérieur à N et qui était devenu hors champ au millésime N) ;
- des logements changent de statut : des RP deviennent RNP et vice-versa.

Pour constituer les grappes à enquêter du T3 2019 au T2 2020, il faut donc faire une étape de mise à jour entre les millésimes 2016 et 2018 qui inclut à la fois le rattachement des nouveaux logements et la mise à jour du statut RP/RNP de chaque logement. (Les grappes enquêtées les années suivantes feront également l'objet de mises à jour annuelles par rapport au millésime précédent.)

Les RNP sont affectées au secteur dont le barycentre des RP est le plus proche, à hauteur d'un seuil limite fixé à 120 dans la plupart des cas (400 pour des communes où la part de RNP est très importante, comme par exemple les stations de sport d'hiver). Au delà du seuil, les RNP

1. Parfois par 7, et dans ce cas un tirage aléatoire simple détermine les 6 grappes choisies pour l'EEC.

les moins proches sont redistribuées en cascade sur les secteurs voisins. Au sein d'un secteur, chaque RNP est alors affectée à la « grappe » parmi les 6 dont le barycentre des RP est le plus proche.

Pour les logements neufs, les volumes étant bien plus faibles, le rattachement s'effectue au secteur le plus proche sans utilisation de seuil, puis au sein du secteur à la grappe la plus proche.

Les grappes obtenues après rattachement des RNP 2016 et des logements neufs 2017 et 2018 sont alors généralement trop grandes pour entrer entièrement en collecte. On les appelle ainsi « grappes-univers ». Pour l'échantillon total de secteurs de l'EEC (principal et complémentaire), on compte 555 994 logements (dont 451 125 RP, soit 81,13 %) répartis en 413 900 étages. Les nouveaux logements sont au nombre de 14 677, soit 2,63 %.

Ces grappe-univers ont en moyenne 25,11 logements. La distribution donne une médiane à 23, et les quartiles et pourcentiles suivants : P1=18, Q1=20, Q3=26, P99=67. Le minimum est 2 (beaucoup de logements disparus à cet endroit), le maximum est 754 (beaucoup de RNP). En termes de type de logement, on a en moyenne 20,38 RP : entre 1 et 242, ce nombre élevé correspondant à une zone ou de nombreux logements qui étaient des RNP en 2016 sont devenus des RP en 2018.

Ces extrêmes indiquent que les grappes-univers ne peuvent pas être directement utilisées en collecte, et qu'un tirage de grappes plus petites doit y être effectué.

2 Tirage des grappes à enquêter au sein des grappes-univers

2.1 Spécification des grappes à tirer

La Division Emploi de l'Insee a précisé les volumes de RP et de RNP devant être présents dans les grappes mises en collecte : en grande majorité 20 RP et 4 à 6 RNP et 95 % des grappes entre 18 et 30 logements, de façon à ne pas faire moins bien que l'EEC actuelle (en moyenne sur 2011 à 2018 : 25,8 logements par grappe, dont 4,4 RNP, et où moins de 10 % des grappes dépassent les 10 RNP). Elle a également choisi les valeurs suivantes des seuils.

Pour les RP :

- Si moins de 24 RP dans la grappe-univers, les prendre toutes
- Sinon en prendre 24

Pour les RNP, on reconduit les seuils de l'EEC actuelle :

- Si moins de 10 RNP dans la grappe-univers, les prendre toutes
- Entre 11 et 40 RNP, en prendre 10
- Entre 41 et 100, en prendre le quart
- Plus de 100, en prendre 25

Une autre contrainte importante pour la collecte est d'enquêter tous les logements d'un même étage. On est ainsi amené à procéder à un tirage d'étages² plutôt que directement de logements. On distingue alors deux types d'étages :

- type A : si une RP au moins est présente (RP au sens 2018)
- type B : sinon

Le tirage d'étages doit permettre de respecter les contraintes sur la distribution de taille des grappes tirées.

2. Il s'agit donc, techniquement parlant, d'un sondage par grappe.

2.2 Étapes du tirage

Trois type de tirages sont a priori possibles : Poissonnien, équilibré, stratifié. Les deux premiers sont écartés pour les raisons suivantes :

- le tirage Poissonnien consiste à donner dans chaque grappe univers une probabilité d'inclusion aux étages de type A d'une part et de type B d'autre part, puis à réaliser un tirage indépendant. Ceci ne convient pas car on ne maîtrise pas la variance i.e. il peut y avoir des grappes quasi-vides ou comprenant quasiment toute la grappe-univers.
- le tirage équilibré sur les contraintes de nombre de RP et de RNP présentes permettrait d'assurer une forte adéquation de la structure des grappes obtenues avec la spécification, mais les probabilités d'inclusion double des étages deviennent alors fortement interdépendantes : certaines configurations seraient systématiquement très peu probables, et l'effet de ceci sur les variables d'intérêt est mal appréhendé.

En définitive, on choisit un tirage aléatoire simple stratifié selon les deux types d'étages.

2.2.1 Première phase

Le calcul des taux de sondage (i.e. des probabilités d'inclusion des étages) ne fait que traduire les spécifications de la Division Emploi. On se place dans une grappe-univers, et on note N_{RP} le nombre de RP présentes. Alors on a :

$$\pi_A = \begin{cases} 1 & \text{si } N_{RP} \leq 24 \\ \frac{24}{N_{RP}} & \text{sinon} \end{cases}$$

Comme les étages de type A contiennent aussi des RNP, ce tirage en apporte en moyenne $n_{RNP}^A = \pi_A N_{RNP}^A = \frac{24}{\max(24, N_{RP})} N_{RNP}^A$, où N_{RNP}^A est le nombre de RNP dans les étages de type A de la grappe-univers.

Le taux de sondage à appliquer à la strate B est ensuite calculé pour :

- respecter en moyenne la cible de nombre de RNP dans l'échantillon
- garantir que ce taux de sondage ne soit pas trop bas, en particulier inférieur à $\frac{1}{N_{RNP}^B}$, pour éviter tout biais (et aussi limiter la dispersion des poids).

On a ainsi :

$$\pi_B = \max\left(\frac{n_{RNP} - n_{RNP}^A}{N_{RNP}^B}, \frac{1}{N_{RNP}^B}\right)$$

Ensuite, les allocations n_A et n_B de nombre d'étages à tirer de chacune des deux strates sont déterminées en arrondissant inférieurement les allocations brutes obtenues avec les taux π_A et π_B , tout en imposant de tirer au moins un étage dans chaque strate pour éviter un biais : $n_A = \max(\lfloor \pi_A N_A \rfloor, 1)$ et $n_B = \max(\lfloor \pi_B N_B \rfloor, 1)$.

2.2.2 Seconde phase

Dans le cas des grappes qui font plus de 40 logements à l'issue de la première phase, une seconde phase est réalisée afin de permettre de ramener la taille en dessous de 41 logements. Ceci n'est pas toujours possible, car il existe des cas où en première phase un étage de plus de 40 logements a été tiré.

Pour le tirage du T3 2019, la seconde phase est réalisée par un tirage réjectif, en définissant de nouvelles cibles en terme d'étages à tirer dans chaque strate. Cette procédure est cependant

problématique³, et n'est pas conservée pour les tirages des trimestres suivants : à la place, une seconde phase par tirage aléatoire simple stratifié est mise en œuvre.

On se place dans une grappe-univers contenant N_A étages de type A (resp. N_B de type B), où un échantillon de première phase S contenant $n_L > 40$ logements a été tiré. On note l'écart à la cible $e := n_L - 40$. S contient n_A étages de type A (resp. n_B de type B), donc les probabilités d'inclusion selon le type d'étage sont $\frac{n_A}{N_A}$ (resp. $\frac{n_B}{N_B}$). On note m_A (resp. m_B) le nombre moyen de logements dans les étages de type A (resp. B) de S .

On cherche alors à déterminer des nombres x_A et x_B d'étages de chaque type à enlever à S . Ceux-ci vont être déterminés par minimisation d'un objectif sous contraintes. Ces dernières sont :

- laisser au moins un étage de type A après seconde phase : $0 \leq x_A < n_A$
- laisser au moins un étage de type B après seconde phase : $0 \leq x_B < n_B$
- x_A et x_B doivent être suffisamment grands pour garantir que l'échantillon de seconde phase contient le plus souvent possible moins de 41 logements

Si la contrainte sur le nombre de logements est difficile à vérifier systématiquement, elle peut néanmoins se formuler en terme d'espérance : $x_A m_A + x_B m_B = e$. Trois possibilités de plan de sondage ont été étudiées : nous présentons ici celle retenue.

On minimise la variance additionnelle générée par le tirage de seconde phase. Conditionnellement à l'échantillon de première phase, cette variance est celle générée par un sondage aléatoire simple stratifié de $n_A - x_A$ étages dans la strate A (resp. $n_B - x_B$ dans la strate B) :

$$V = n_A^2 \left(1 - \frac{n_A - x_A}{n_A}\right) \frac{s_A^2}{n_A - x_A} + n_B^2 \left(1 - \frac{n_B - x_B}{n_B}\right) \frac{s_B^2}{n_B - x_B}$$

La contrainte est qu'en espérance les étages enlevés cumulent e logements : $x_A m_A + x_B m_B = e$. On trouve alors des allocations de Neyman avec contraintes de coût (voir Tillé [6] page 81) :

$$\begin{cases} x_A = n_A - \frac{n_A}{\sqrt{m_A}} \frac{n_A m_A + n_B m_B - e}{n_A \sqrt{m_A} + n_B \sqrt{m_B}} \\ x_B = n_B - \frac{n_B}{\sqrt{m_B}} \frac{n_A m_A + n_B m_B - e}{n_A \sqrt{m_A} + n_B \sqrt{m_B}} \end{cases}$$

Enfin ces allocations théoriques doivent être arrondies. On choisit de le faire à l'entier immédiatement supérieur⁴, sous contrainte de tirer au moins un étage dans chaque strate.

2.2.3 Pondérations

La variable de poids associée à toutes étapes de construction des grappes en collecte permet d'effectuer sur un échantillon pour un trimestre donné, par exemple le T3 2019, une estimation sans biais des variables de population. Ce poids est le produit :

- du poids des unités de coordination obtenu par application de la méthode généralisée du partage des poids
- de l'inverse de la probabilité de sélection utilisée pour le tirage des secteurs conditionnellement aux unités de coordination

3. Les probabilités de sélection sont alors inconnues. Seules des simulations coûteuses en temps de calcul pourraient les déterminer. De plus, ce tirage a tendance à favoriser les étages ayant peu de logements car ils permettent de mieux satisfaire les contraintes. Certains étages pourraient même avoir des probabilités nulles d'apparaître, ce qui entraînerait un biais de défaut de couverture. Tous ces aspects rendent le tirage réjectif inadapté.

4. Cela conduit à retirer plus de logements, et permet de diminuer la probabilité que l'échantillon dans cette grappe-univers contienne encore 41 logements ou plus.

- d'un facteur égal à 6, tenant compte de l'affectation du secteur l'un des six trimestres d'entrée
- d'un facteur égal à (nombre de grappes du secteur) / 6 . Ce facteur permet de tenir compte du fait que, dans les secteurs contenant 7 grappes, seules 6 grappes, sélectionnées aléatoirement par un sondage aléatoire simple, sont mises en collecte
- d'un facteur égal à 6, tenant compte de la sélection de la grappe interrogée dans le secteur
- du poids conditionnel des étages dans les grappes

Et, finalement, le poids des logements est égal au poids de leur étage, puisque l'on prend tous les logements d'un étage tiré.

2.3 Résultats pour les tirages du T3 2019 au T2 2020

2.3.1 Nombre de logements des grappes obtenues

Les 4 tirages effectués pour construire les grappes entrants aux trimestres T3 2019 à T4 2020 présentent des propriétés tout à fait satisfaisante vis-à-vis des spécifications. Elles sont notamment meilleures que celles de l'EEC actuelle sur la période 2011-2018. Ceci est résumé dans le tableau suivant, où l'on sépare le T3 2019 au trois autres trimestres étant donné la différence du traitement de la seconde phase pour les grappes dépassant initialement les 40 logements.

Nombre de logements par grappe (%)	EEC actuelle	Tirage T3 2019	Tirage T4 2019 + T1 2020 + T2 2020
Moins de 4	0,1	0,6	0,1
Entre 5 et 9	0,1	0,0	0,1
Entre 10 et 14	0,2	0,6	0,4
Entre 15 et 19	3,7	7,5	5,4
Entre 20 et 24	43,7	59,2	62,1
Entre 25 et 26	13,7	11,0	10,2
Entre 27 et 29	15,7	12,4	11,5
Entre 30 et 34	16,3	7,5	8,3
Entre 35 et 39	5,6	0,8	1,4
Entre 40 et 44	0,8	0,4	0,4
Plus de 44	0,2	0,0	0,1
Total	100,0 [16 595 grappes]	100,0 [493 grappes]	100,0 [1 846 grappes]
Nombre moyen de RP	21,4	19,8	19,9
Nombre moyen de RNP	4,4	3,5	3,6
Pourcentile P90	33	29	30

TABLE 1 – Comparaison des nombres de logements des grappes en collecte de l'EEC actuelle et des tirages pour la nouvelle EEC Nautile

Par ailleurs, la dispersion des poids est stable sur les 4 tirages, et comparable à celle de l'EEC actuelle. En effet, le rapport de pourcentiles P99/P1 vaut 5,1 pour l'EEC actuelle, et entre 4,6 et 6,4 pour ces tirages Nautile.

2.3.2 Caractéristiques terrain des grappes obtenues

Les longueurs terrain des grappes en collecte sont meilleures que celles de l'EEC actuelle dans toutes les régions, que ce soit au niveau moyen ou médian. Pour le haut de la distribution (pourcentile P99), il peut arriver d'avoir une situation moins bonne, de manière ponctuelle sur un trimestre dans des zones urbaines. Les zones rurales sont en particulier plus compactes. Ceci est synthétisé dans le tableau suivant, où l'on ne donne les résultats, outre en France métropolitaine, que pour trois régions assez caractéristiques.

Longueur (en km) [et nombre grappes]	France métropolitaine	Ile-de-France	Aquitaine	Corse
moyenne T4 2018 à T2 2019	2,9 [1607]	0,8 [285]	4,2 [80]	10,2 [9]
moyenne T3 2019 Nautile (P)	2,2 [493]	0,7 [93]	3,8 [29]	2,9 [6]
moyenne T4 2019 Nautile (P+S)	2,3 [618]	0,4 [119]	3,2 [30]	8,2 [7]
moyenne T1 2020 Nautile (P+S)	2,4 [621]	0,8 [112]	3,3 [30]	5,6 [6]
moyenne T2 2020 Nautile (P+S)	2,1 [607]	0,5 [116]	2,8 [28]	3,9 [8]
médiane T4 2018 à T2 2019	1,4 [1607]	0,8 [285]	2,4 [80]	0,4 [9]
médiane T3 2019 Nautile (P)	0,9 [493]	0,2 [93]	2,0 [29]	1,6 [6]
médiane T4 2019 Nautile (P+S)	0,9 [618]	0,0 [119]	1,3 [30]	3,4 [7]
médiane T1 2020 Nautile (P+S)	1,0 [621]	0,2 [112]	1,9 [30]	0,1 [6]
médiane T2 2020 Nautile (P+S)	0,8 [607]	0,2 [116]	1,5 [28]	2,0 [8]
P99 T4 2018 à T2 2019	19,9 [1607]	7,3 [285]	32,2 [80]	37,1 [9]
P99 T3 2019 Nautile (P)	16,5 [493]	10,5 [93]	15,3 [29]	9,2 [6]
P99 T4 2019 Nautile (P+S)	23,8 [618]	4,0 [119]	27,7 [30]	28,8 [7]
P99 T1 2020 Nautile (P+S)	19,8 [621]	5,1 [112]	18,1 [30]	19,8 [6]
P99 T2 2020 Nautile (P+S)	20,4 [607]	3,0 [116]	24,7 [28]	12,9 [8]

TABLE 2 – Comparaison avec Metric de la longueur des grappes de l’EEC actuelle et de la nouvelle EEC Nautile

2.3.3 Premiers résultats

Les premiers chiffres produits par la nouvelle EEC en utilisant une collecte sur les grappes tirées pour le T3 2019 ont été publiés le 14 novembre 2019 sur [Insee.fr](https://www.insee.fr)⁵. En particulier, ces nouvelles grappes ont donné satisfaction, tant au niveau de la collecte terrain, que sur le calage.

Bibliographie

[1] Costa, L. et Guillo, C. et Paliot, N. et Merly-Alpa, T. et Vincent, L. et Chevalier, M. et Deroyon, T. (2018), « Le Tirage Coordonné du Nouvel Échantillon-Maître Nautile avec l’Échantillon de l’Enquête Emploi en Continu », Actes des JMS 2018.

[2] Guillaumat-Tailliet F., Tavan C., « Une nouvelle enquête Emploi en 2021. Entre impératif européen et volonté de modernisation », *Courrier des statistiques* n° 6, Insee, juillet 2021.

[3] Loonis, V., « La Construction du Nouvel Échantillon de l’Enquête Emploi en Continu à Partir des Fichiers de la Taxe d’Habitation », Actes des JMS 2009.

[4] Merly-Alpa, T. et Vincent, L. et Costa, L. et Guillo, C. et Paliot, N. et Chevalier, M. et Deroyon, T. (2018), « Tirage Coordonné d’Échantillons : une Application à l’Échantillon-Maître Nautile et à l’Enquête Emploi », Actes du 10e Colloque Francophone sur les Sondages 2018.

[5] Quénechdu, V. et Tchobanian; C. (2021), « L’enquête Emploi se rénove en 2021 : des raisons de sa refonte aux impacts sur la mesure de l’emploi et du chômage », *Insee Analyses*, n°65, juin 2021.

[6] Tillé, Y., *Théorie des Sondages* (2e édition), Dunod.

5. Voir <https://www.insee.fr/fr/statistiques/4247277>