

REDRESSEMENT DE L'ENQUÊTE EMPLOI ANNUELLE DE MAYOTTE 2023

Joachim Timotéo ¹ & Christian Monteil ²

¹ *Insee, France, joachim.timoteo@insee.fr*

² *Insee, France, christian.monteil@insee.fr*

Résumé. La collecte de l'enquête emploi annuelle de Mayotte 2023 a été perturbée par l'opération Wuambushu. Il s'agit d'une opération policière française en cours à Mayotte depuis le 24 avril 2023, visant à expulser les étrangers en situation irrégulière, à détruire les bidonvilles et à lutter contre la criminalité dans l'archipel. Afin de garantir la sécurité des enquêteurs, il a été décidé d'arrêter la collecte à partir du 24 avril dans les zones à fort risques d'agression en lien avec l'opération.

En amont des opérations de correction de la non-réponse par repondération avec groupes homogènes de réponse et de calage sur marges, le biais de collecte a été étudié donnant lieu à plusieurs modèles d'estimation des probabilités de réponse à l'enquête. Une hypothèse a été ajoutée dans les marges de calage, consistant à reconduire en 2023 la répartition des 15 ans ou plus selon leur lieu de naissance de 2022.

Mots-clés. Traitement de la non réponse, taux de réponse faibles et biais de sélection, estimations et calage

Abstract. The annual employment survey of Mayotte in 2023 was disrupted by the Wuambushu operation. This French police operation occurred in Mayotte since April 24, 2023. It aimed at expelling illegal foreigners, destroying shanty towns and fighting crime in the archipelago. In order to guarantee the safety of pollsters, it was decided to stop the collection from April 24 in areas with a high risk of aggression linked to the operation.

Before the non-response treatment operations by reweighting with homogeneous response groups and calibration on margins, the collection bias was studied with several models for the estimation of the response probabilities to the survey. An hypothesis was added to the calibration margins, consisting of renewing in 2023 the distribution of the population over 15 years old according to the place of birth in 2022.

Keywords. Non-response treatment, low response rates and selection bias, estimation and calibration

1 Contexte de l'enquête emploi annuelle de Mayotte 2023

1.1 L'échantillonnage de l'enquête

L'échantillon de l'enquête emploi annuelle (EEA) à Mayotte est un panel rotatif d'environ 3000 logements où chaque logement est enquêté trois années consécutives avant de sortir de l'échantillon. L'échantillon est donc renouvelé chaque année d'un tiers.

La base de sondage de l'EEA correspond à l'enquête cartographique de 2022 (groupe de rotation 5). Il s'agit d'une base d'adresses situées dans l'ensemble des 17 communes du département. Il n'y a aucune exclusion d'îlots pour cause d'accès difficile. Le champ statistique recouvre les adresses ayant un nombre positif de logements qualifiés d'ordinaires.

Le tirage des adresses entrantes de l'enquête est équilibré selon la méthode du cube [Deville, J.-C. et Tillé, Y. (2004)] sur quatre variables et stratifié par secteur d'action enquêteur (SAE) afin de répartir la charge entre les enquêteurs. Les 26 SAE de Mayotte ont été élaborés en 2021 par agrégation d'îlots selon l'algorithme du voyageur de commerce, chaque zone ayant in fine un volume de logements comparable [Jamme J. (2021)].

Les quatre variables d'équilibrage sont le type d'adresse (mono-logement, 2 à 4 logements, plus de 4), le type de bâti (en dur ou non), un regroupement géographique des communes (8 groupes) et une classification des villages en 4 groupes selon leur situation socio-économique [Thibault, P. (2019)].

L'année 2023 est le dernier millésime de l'EEA. Depuis le premier trimestre 2024, Mayotte a rejoint le dispositif de l'enquête emploi en continu, dont la méthodologie n'est pas présentée ici.

1.2 Un protocole perturbé avec un fort risque de biais de collecte

La collecte de l'EEA 2023 a été perturbée par l'opération Wuambushu. Il s'agit d'une opération policière française en cours à Mayotte depuis le 24 avril 2023, visant à expulser les étrangers en situation irrégulière, à détruire les bidonvilles et à lutter contre la criminalité dans l'archipel. Afin de garantir la sécurité des enquêteurs, il a été décidé d'arrêter la collecte à partir du 24 avril dans les zones à fort risques d'agression en lien avec l'opération. Les logements non enquêtés dans ces zones ont été considérés hors champ par la maîtrise d'ouvrage pour des raisons de suivi d'enquête.

L'inaccessibilité de certaines zones a non-seulement fait chuter le nombre de répondants mais également fait peser un fort risque de biais sur les résultats de collecte dans la mesure où les caractéristiques socio-démographiques de ces zones sont sensiblement différentes du reste du territoire. Dans ces zones, la part des nés à l'étranger est significativement plus forte et il en est de même pour la part d'inactifs. De plus, ces caractéristiques sont assez fortement corrélées avec l'emploi et ont sensiblement évolué entre 2022 et 2023. Ainsi, par rapport à la précédente enquête, la proportion des répondants de moins de 30 ans s'est accrue dans les

zones Wuambushu, contrairement à la part des nés à l'étranger, significativement plus faible.

2 Traitement de la non-réponse totale

La méthode de traitement de la non-réponse totale a été adaptée suite aux analyses sur le biais de collecte lié aux changements de protocole. Néanmoins la méthode mobilisée reste identique dans ses grandes lignes à savoir la méthode par repondération des logements [Ardilly, P. (2006)] en procédant tout d'abord à une correction de la non-réponse au niveau logement puis au niveau individuel.

2.1 Traitement de la non-réponse au niveau logement : estimation des probabilités de réponse et groupes de réponses homogènes

La base de sondage pour les enquêtes à Mayotte étant de niveau adresse, le nombre et la nature des variables explicatives de la non-réponse sont limités. Ainsi, les covariables retenues dans le modèle de régression logistique sont :

- le type de village d'appartenance (G1 = villages cumulant les difficultés, G2 = villages aux conditions de logement précaires, G3 = villages aux conditions de vie supérieures à la moyenne, G4 = villages aux conditions de vie les plus favorables) [Thibault, P. (2019)]
- le type de bâti (fortune, bois ou en dur)
- le regroupement de communes d'appartenance (8 groupes)
- l'indicatrice de présence dans un "quartier politique de la ville" (zones à très faibles revenus)
- le type d'adresses du logement (mono-logement, 2 à 4 logements, plus de 4)
- l'indicatrice de présence en zone Wuambushu

Le modèle retenu permet ainsi d'expliquer 92% de la déviance du modèle nul, c'est-à-dire sans aucun prédicteur [Figure 1].

À partir de ce modèle, cinq groupes de réponses homogènes (GRH) ont été constitués en utilisant la méthode proposée par [Haziza, D. et Beaumont, J.-F. (2007)] reposant sur l'utilisation de l'algorithme des centres-mobiles (ou *k-means*). La probabilité de réponse moyenne varie de 59.2% à 92.7% selon les groupes. Ainsi le facteur de correction de la non-réponse, appliqué aux poids de sondage, varie entre 1.1 et 1.7.

De plus, le poids des logements appartenant à une zone Wuambushu a été dilaté par le rapport entre le nombre de logements répondants dans ces zones en 2022 et celui de 2023. Cette repondération vise à compenser la mise en hors champ des logements en zone

| | | Estimate | Std. Error | z value | Pr(> z) | |
|----------------|---------------------|----------|------------|---------|----------|-----|
| | (Intercept) | 1,617 | 0,280 | 5,776 | 0,000 | *** |
| groupe_village | G1 | ref | ref | ref | ref | |
| | G2 | -0,209 | 0,214 | -0,979 | 0,327 | |
| | G3 | -0,272 | 0,235 | -1,157 | 0,247 | |
| | G4 | -0,736 | 0,251 | -2,935 | 0,003 | ** |
| type_bati | fortune | ref | ref | ref | ref | |
| | bois | -1,452 | 0,640 | -2,270 | 0,023 | * |
| | dur | -0,978 | 0,148 | -6,595 | 0,000 | *** |
| groupe_com | G1 | ref | ref | ref | ref | |
| | G2 | 0,956 | 0,274 | 3,489 | 0,000 | *** |
| | G3 | 0,523 | 0,226 | 2,317 | 0,021 | * |
| | G4 | 0,915 | 0,235 | 3,886 | 0,000 | *** |
| | G5 | 0,214 | 0,248 | 0,864 | 0,388 | |
| | G6 | 0,792 | 0,263 | 3,016 | 0,003 | ** |
| | G7 | 0,640 | 0,253 | 2,535 | 0,011 | * |
| | G8 | 0,117 | 0,213 | 0,550 | 0,582 | |
| i_qpv | 0 | ref | ref | ref | ref | |
| | 1 | 0,189 | 0,130 | 1,457 | 0,145 | |
| type_adr | mono_logement | ref | ref | ref | ref | |
| | 2 à 4 logements | -0,072 | 0,126 | -0,573 | 0,567 | |
| | plus de 4 logements | -1,402 | 0,310 | -4,520 | 0,000 | *** |
| i_Wuam | 0 | ref | ref | ref | ref | |
| | 1 | 1,463 | 0,501 | 2,918 | 0,004 | ** |

Signif. Codes : 0 '***' ; 0.001 '**' ; 0.01 '*' ; 0.05 '.' ; 0.1 '.' ; 1

Null deviance : 2307
Residual deviance : 2118
AIC : 2152
Pseudo R² 0,92

Figure 1: modèle d'estimation de la non-réponse au niveau logement

Wuambushu et de les considérer plutôt comme des non-répondants. Cela revient à considérer une taille d'échantillon en zone Wuambushu en 2023 égale à celle de 2022 [équations 1 et 2], ainsi que des caractéristiques semblables entre répondants et non répondants en zone Wuambushu en 2023.

En effet, l'estimation du total de logements en zone Wuambushu en 2023 (\hat{N}^{W23}) peut s'écrire comme la somme des poids dans le sous-échantillon des logements répondants en zone Wuambushu (S_r^W):

$$\begin{aligned}\hat{N}^{W23} &= \sum_{k \in S_r^W} \frac{1}{\pi_k \cdot \hat{p}_k} \\ &= \frac{N^{23}}{n^{23}} \cdot n^{W23}\end{aligned}\quad (1)$$

avec :

$\pi_k = \frac{n^{23}}{N^{23}}$ la probabilité d'inclusion d'un logement k en 2023 où n^{23} est la taille de l'échantillon

et N^{23} le nombre total de logements

\hat{p}_k la probabilité de réponse d'un logement k

n^{W23} la taille de l'échantillon en zone Wuambushu en 2023

Or, cette estimation souffre d'un défaut de couverture du fait de l'opération militaire. Faute d'informations supplémentaires, les sous-échantillons de répondants 2022 et 2023 ont été supposés proportionnels. C'est pourquoi, en pondérant (1) par le rapport entre le nombre de répondants en zone Wuambushu en 2022 et celui de 2023 ($\frac{r^{W22}}{r^{W23}}$), la nouvelle estimation du total de logements en zone Wuambushu en 2023 (\hat{N}^{*W23}) devient :

$$\begin{aligned}\hat{N}^{*W23} &= \frac{N^{23}}{n^{23}} \cdot n^{W23} \cdot \frac{r^{W22}}{r^{W23}} \\ &= \frac{N^{23}}{n^{23}} \cdot n^{W22}\end{aligned}\quad (2)$$

Ainsi, l'estimation du nombre total de logements à Mayotte en 2023 (\hat{N}^{23}), après traitement de la non-réponse au niveau logement, peut s'écrire comme la somme de deux totaux sur deux sous-échantillons de répondants, l'un hors zone Wuambushu (S_r^{HW}), l'autre dans la zone (S_r^W):

$$\begin{aligned}\hat{N}^{23} &= \sum_{k \in S_r^{HW}} \frac{1}{\pi_k \cdot \hat{p}_k} + \sum_{k \in S_r^W} \frac{1}{\pi_k \cdot \hat{p}_k} \cdot \alpha \\ &= \frac{N^{23}}{n^{23}} \cdot n^{HW23} + \frac{N^{23}}{n^{23}} \cdot n^{W22} \\ &= \hat{N}^{HW23} + \hat{N}^{*W23}\end{aligned}\quad (3)$$

avec :

$\pi_k = \frac{n^{23}}{N^{23}}$, probabilité d'inclusion d'un logement k où n^{23} est la taille de l'échantillon et N^{23} est le nombre total de logements dans la base de sondage

\hat{p}_k la probabilité de réponse d'un logement k

$\alpha = \frac{r^{W22}}{r^{W23}}$ rapport entre le nombre de répondants dans les zones Wuambushu en 2022 et celui de 2023 (hypothèse de proportionnalité des sous-échantillons).

2.2 Traitement de la non-réponse au niveau individu

Le traitement de la non-réponse individuelle n'est pas concernée par l'opération Wuambushu et suit la procédé habituel visant à corriger le biais de surestimation lié à la chaîne aval de la collecte. En effet, dans le traitement des données de collecte, tous les individus non répondants vivant dans un logement répondant sont conservés et imputés par défaut en inactifs. Pour corriger ce biais, nous supprimons donc ces individus non répondants et dilatons le poids des individus répondants habitant dans le même logement.

Ainsi, en repartant de (3), l'estimation du total de la population ($p\hat{op}^{23}$), après traitement de la non-réponse au niveau individu, peut de nouveau s'écrire comme la somme de totaux

sur deux sous-échantillons de répondants, l'un hors zone Wuambushu (S_r^{HW}), l'autre dans la zone (S_r^W) :

$$\begin{aligned} \hat{p}p^{23} &= \sum_{k \in S_r^{HW}} \frac{1}{\pi_k \cdot \hat{p}_k} \cdot \frac{\sum_{i \in k | S_r^{HW}} 1_i}{\sum_{i_r \in k | S_r^{HW}} 1_{i_r}} + \sum_{k \in S_r^W} \frac{1}{\pi_k \cdot \hat{p}_k} \cdot \alpha \cdot \frac{\sum_{i \in k | S_r^W} 1_i}{\sum_{i_r \in k | S_r^W} 1_{i_r}} \\ &= \hat{p}p^{HW23} + \hat{p}p^{*W23} \end{aligned} \quad (4)$$

avec :

1_i l'indicatrice de présence d'un individu i dans un logement répondant

1_{i_r} l'indicatrice de présence d'un individu répondant i_r dans un logement répondant.

3 Calage sur marges

À l'issue de la correction de la non-réponse, l'étape de calage permet de retrouver la cohérence en niveau avec des totaux de référence, ainsi que d'améliorer la précision des estimations.

Les marges classiquement retenues sont la pyramide des âges, le nombre de résidences principales et la population par lieu de naissance. Ces deux dernières sont issues d'un calcul effectué par le Pôle d'ingénierie des enquêtes ménages de l'Insee spécifiquement pour le calage sur marges. Les marges officielles restent à ce jour celles du Recensement de 2017.

Afin de contrôler au maximum le biais de collecte, trois autres marges ont été ajoutées cette année : le nombre de mono-logements, de bâtis en dur ainsi que le nombre de logements dans un quartier politique de la ville. Ces variables sont fortement corrélées aux variables d'intérêt. Elles interviennent pour certaines non seulement dans le plan de sondage mais aussi dans le modèle de correction de la non-réponse totale.

La **pyramide des âges** est fournie par la Direction de l'action régionale (DDAR) de l'Insee qui publie des estimations de population mahoraise par sexe et âge au 1er janvier. Cette marge se décompose en 13 modalités pour chaque sexe.

Le **nombre de résidences principales** en 2023 est la moyenne de deux estimations, l'une issue du Recensement de la population ($r\hat{e}sp_{2023}^{rp}$), l'autre du Répertoire d'immeubles localisés ($r\hat{e}sp_{2023}^{ril}$) :

$$r\hat{e}sp_{2023} = \frac{r\hat{e}sp_{2023}^{rp} + r\hat{e}sp_{2023}^{ril}}{2} \quad (5)$$

Notations :

$resp_{gr,N}^{ril}$, le nombre de résidences principales dans le groupe de rotation du RIL de l'année N

$duree_{2017a2023}$, le nombre de jours entre le RP 2017 et le 1er janvier 2023

$duree_{2012a2017}^{rp}$, le nombre de jours entre les RP 2012 et 2017

$duree_{2017aN}^{ril}$, le nombre de jours entre le RIL 2017 et le RIL de l'année N

$nombreg_r$, le nombre de groupes de rotation depuis 2017 (ici : 6)

$r\hat{e}sp_{2023}^{rp}$ est l'estimation du nombre de résidences principales en 2023 à partir des RP 2012 et

2017. Il s'agit de la prolongation de tendance du nombre de résidences principales mesurée entre les deux derniers recensements :

$$r\hat{e}sp_{2023}^{rp} = resp_{2017}^{rp} \cdot \left(\frac{resp_{2017}^{rp}}{resp_{2012}^{rp}} \right)^{\frac{duree_{2017a2023}^{rp}}{duree_{2012a2017}^{rp}}} \quad (6)$$

$r\hat{e}sp_{2023}^{ril}$ est l'estimation du nombre de résidences principales à partir des groupes de rotations des RIL 2017 à 2023. Il s'agit de la projection du nombre de résidences principales au RP 2017 selon l'évolution moyenne dans le RIL entre 2017 et 2023.

$$r\hat{e}sp_{2023}^{ril} = \frac{resp_{2017}^{rp}}{nombregr} \cdot \sum_{gr=1, N=2018}^{gr=6, N=2023} \left(\frac{resp_{gr, N}^{ril}}{resp_{gr, 2017}^{ril}} \right)^{\frac{duree_{2017a2023}^{ril}}{duree_{2017aN}^{ril}}} \quad (7)$$

Le **lieu de naissance des individus âgés de 15 ans ou plus** est nécessaire au calage de l'enquête du fait de sa très forte corrélation avec l'emploi. Il est décliné en trois modalités (Mayotte, France hors Mayotte, étranger). Sa dernière mesure datant du RP 2017, cette marge est calculée à partir de l'estimation de population au 1er janvier de la DDAR à laquelle est appliquée la dernière répartition connue au RP 2017, actualisée par l'évolution mesurée dans les résultats de l'EEA elle-même [(8)].

La répartition par lieu de naissance en 2023 a été exceptionnellement supposée identique à celle observée en 2022 en raison d'une évolution significativement forte liée à l'opération Wuambushu, incohérente avec les dires d'experts.

$$p\hat{op}_{ln}^{23} = estimPop_{1erjanv23} \times part_{ln}^{rp17} \cdot \frac{part_{ln}^{eea22}}{part_{ln}^{eea17}} \times \frac{100}{\sum_{ln} part_{ln}^{rp17} \cdot \frac{part_{ln}^{eea22}}{part_{ln}^{eea17}}} \quad (8)$$

4 Quelques résultats provisoires

En 2023, la part de chômeurs augmente de 1,6 point entre les poids de tirage et les poids finaux redressés et calés. Le taux de chômage s'accroît de 3 points après post-traitements pour s'établir ainsi à 36,4% contre 34,5% en 2022 [Figure 2].

Il convient toutefois de relativiser cette hausse du fait de l'importance du choix de garder la répartition des lieux de naissance telle que calculée en 2022. En effet, si nous avons appliqué la répartition observée en 2023, le taux de chômage aurait été sensiblement égal à 35%.

| 2023 | Poids initiaux | Poids corrigés de la non-réponse | Poids finaux |
|-----------------|----------------|----------------------------------|--------------|
| | En niveau | | |
| En emploi | 24 375 | 34 719 | 50 908 |
| Au chômage | 12 244 | 16 624 | 29 093 |
| Inactif | 48 658 | 67 090 | 102 048 |
| Ensemble | 85 276 | 118 434 | 182 048 |
| | En % | | |
| En emploi | 28,6 | 29,3 | 28,0 |
| Au chômage | 14,4 | 14,0 | 16,0 |
| Inactif | 57,1 | 56,6 | 56,1 |
| Ensemble | 100,0 | 100,0 | 100,0 |
| Taux de chômage | 33,4 | 32,4 | 36,4 |
| Taux d'emploi | 28,6 | 29,3 | 28,0 |

Source : Enquête emploi Mayotte 2023
Champ : résidents mahorais de 15 ans ou plus au 31 décembre 2023 (age3112 ≥ 15)

Figure 2: Situation sur le marché du travail à Mayotte avant et après redressements

Bibliographie

- Deville, J.-C. et Tillé, Y. (2004), *Efficient Balanced Sampling : The Cube Method*, ed. Biometrika, Vol 91, No 4, pp 893-912.
- Ardilly, P. (2006), *Les techniques de sondage*, ed. TECHNIP, pp. 370-382.
- Haziza, D. et Beaumont, J.-F. (2007), *On the construction of imputation classes in surveys*, *International Statistical Review*, 75, pp. 25-43.
- Thibault, P. (2019), *Les villages de Mayotte en 2017*, *Insee Analyses Mayotte*, 22.
- Jamme, J. (2021), *Élaboration d'une méthodologie de tirage de l'échantillon de l'Enquête Emploi en Continu à Mayotte et Estimation de la précision par simulations*, *Mémoire de stage Master Ensai*, 2021_24325_DR974-SES.
- Monteil, C. (2023), *Tirage de l'échantillon 2023 de l'Enquête emploi annuelle à Mayotte*, *Note interne Insee*, 2023_4409_DR974-SES.
- Timotéo, J. (2024), *Note de calage de l'Enquête emploi annuelle à Mayotte*, *Note interne Insee*, 2024_10412_DR974-SES.