

Estimation de la précision dans l'enquête Formation tout au long de la vie à Mayotte par utilisation du package *Gustave*

Romain Loiseau¹ & Joachim Timotéo²
¹*Insee, France, romain.loiseau@insee.fr*
²*Insee, France, joachim.timoteo@insee.fr*

Résumé.

L'enquête Formation tout au long de la vie a pour thématique la formation et les compétences des adultes. La collecte réalisée à Mayotte pour l'année 2022 représente 2 000 logements, échantillonnés sur la base de données cartographiques. Cette spécificité nécessite d'être prise en compte dans l'estimation de la précision des indicateurs produits. Depuis la variance liée à l'échantillonnage d'adresses à celle résultant des redressements de l'enquête, la variabilité des différentes étapes du mécanisme de réponse menant aux données d'enquête est considérée afin de mesurer au plus juste l'incertitude inhérente à l'échantillonnage. Le package *gustave*, développé en R et maintenu à l'Insee, offre toute une gamme d'outils facilitant les calculs de précision dans les enquêtes par sondage. Les possibilités que le package propose vont de l'implémentation des estimateurs de variance à la production simple des informations nécessaires aux calculs de précision, dont les intervalles de confiance. Cet outil a donc été employé pour estimer la variance de l'ensemble des indicateurs issus de l'enquête FLV à Mayotte et permettre une diffusion facilitée des résultats auprès de tous les acteurs concernés.

Mots-clés.

Estimation de variance, logiciel d'analyse des données d'enquête.

Abstract.

The Lifelong Learning survey focuses on adult training and skills. In Mayotte, the 2022 survey edition will cover 2,000 homes, sampled on the basis of cartographic data. This specificity needs to be taken into account when estimating the precision of the indicators produced. From the variance linked to address sampling to that resulting from survey adjustments, the variability of the various stages of the response mechanism leading to the survey data is considered in order to measure the overall uncertainty as precisely as possible. The *gustave* package, developed with R software and maintained at INSEE, offers a whole range of tools to facilitate precision calculations in sample surveys. The possibilities that the package offers range from the implementation of variance estimators to an easy production of the information required for precision calculations, including confidence intervals. This tool has been used to estimate the variance of all the indicators from the Lifelong Learning survey in Mayotte, making it easier to share the results with all the stakeholders concerned.

Keywords.

Variance estimation, survey data analysis software.

1 L'enquête FLV et sa spécificité à Mayotte

L'enquête Formation tout au long de la vie (FLV) a pour thématique la formation et les compétences des adultes âgés de 18 à 69 ans vivant en logement ordinaire, tant sur le territoire métropolitain que dans les départements d'Outre-Mer. Elle comporte deux volets principaux. L'un vise à mesurer

l'accès des adultes à la formation, qu'elle soit formelle, non formelle ou qu'il s'agisse d'apprentissage informel. L'autre volet cherche à mesurer les compétences des adultes à l'écrit et en numératie dans diverses situations de vie courante. L'enquête collecte également un certain nombre d'informations socio-démographiques telles que l'âge, le sexe, le niveau de diplôme, ainsi que des informations sur les langues maîtrisées et les origines géographiques. D'autres thématiques sont abordées par l'enquête, comme l'accès à l'information sur les possibilités de formation, la connaissance des droits et les obstacles rencontrés dans l'accès à la formation.

Historiquement, l'enquête FLV a lieu tous les six ans. L'édition 2022 reprend en partie dans son volet compétences des éléments issus de l'enquête Information et vie quotidienne (IVQ), dont la dernière passation date de 2011, afin de reproduire certains indicateurs tels que la mesure de l'illettrisme.

L'enquête FLV concerne le territoire métropolitain ainsi que les cinq départements d'Outre-Mer. Elle s'inscrit dans un contexte plus large que le territoire national, au niveau européen, puisque l'ensemble des pays de l'Union Européenne sont concernés. Par ailleurs, plusieurs indicateurs sont soumis à des contraintes de précision statistique, dont le niveau d'exigence est fixé au niveau européen et dépend du territoire et de la population concernée.

En pratique, l'enquête est menée par la Dares¹ conjointement avec l'Insee, et collecte des données relatives aux années 2022 et 2023. EU effectif total de 30 000 logements a été tiré dont 10 000 dans les cinq départements d'Outre-Mer, à hauteur de 2 000 logements par département. Deux vagues de collecte en face-à-face ont été réalisées successivement et mobilisent des données diverses. Le fichier Fidéli, produit à partir de données fiscales, est couramment utilisé pour échantillonner des logements ou des individus dans les enquêtes ménages à l'Insee, comme ici pour la partie métropolitaine. Le Recensement de la population 2018 a par ailleurs été employé pour le tirage des deux vagues de collecte réalisées en Outre-Mer, hors Mayotte, afin de permettre une représentativité régionale. Les enquêtes cartographiques 2021 et 2022 ont quant-à-elles été mobilisées pour l'une et l'autre des vagues de collecte à Mayotte, et permettent également une telle représentativité.

Le tirage des échantillons de collecte dans des bases de données cartographiques constituent une spécificité courante concernant Mayotte dans le cadre des enquêtes statistiques dont le tirage est réalisé par l'Insee, en raison de la moindre qualité des données provenant de Fidéli et du Recensement de la population sur ce territoire. Les enquêtes cartographiques permettent de constituer des bases de données d'adresses, réparties suivant leur taille. Les mono-logements désignent les adresses auxquelles un seul logement est connu. Les petites adresses désignent les adresses pour lesquelles deux à quatre logements sont recensés. Les grandes adresses désignent les adresses où plus de quatre logements sont recensés.

Le calcul de précision des indicateurs de l'enquête FLV doit de fait prendre en compte cette spécificité et nécessite des hypothèses.

1.1 Description du plan de sondage

L'enquête FLV à Mayotte représente un échantillon de 2 000 logements, répartis en deux vagues de collecte : 1 300 logements enquêtés en 2022 et 700 logements en 2023. Le tirage de l'échantillon de la première vague de collecte a été réalisé sur la base de l'enquête cartographique de 2021, tandis que l'échantillon de la deuxième vague concerne l'enquête cartographique de 2022.

¹ La Direction de l'animation de la recherche, des études et des statistiques (Dares) est la direction statistique du ministère du Travail, du Plein emploi et de l'Insertion.

Aucune restriction de champ, comme l'exclusion d'îlots pour cause d'accès difficile, n'a été appliquée. Néanmoins, les adresses qui ont déjà été tirées pour une enquête au cours des cinq dernières années ont été ôtées de la base de tirage ; ces retraits ne concernant que les adresses mono-logement et les petites adresses.

La constitution de l'échantillon d'enquête correspond à l'application d'un plan de sondage à plusieurs degrés. La base des données cartographiques à Mayotte, elle-même le produit d'une enquête, est considérée comme exhaustive dans ces travaux, et aucun aléa dans les estimations produites n'est donc imputé à cet égard.

Au premier degré, un tirage des adresses est réalisé de manière équilibrée sur chaque secteur d'action enquêteur selon la méthode du cube sur quatre variables : le type d'adresse, le type de bâti, une variable géographique sur le groupe des communes, et une classification des villages. Au deuxième degré, un tirage aléatoire simple de logements est réalisé à chaque grande adresse échantillonnée au premier degré, tandis que l'ensemble des mono-logements et petites adresses sont conservés.

1.2 Redressement des données de collecte

À la suite de la passation des deux vagues d'enquête à Mayotte, un phénomène classique de non-réponse est observé. Compte tenu de celle-ci, des travaux de redressement ont été réalisés. Un traitement de la non-réponse puis un calage au niveau ménage sont effectués au niveau des logements, avant le traitement de la non-réponse au niveau individuel suivi d'un nouveau calage.

La première étape des redressements est donc le traitement de la non-réponse totale au niveau des logements, par repondération. Elle consiste à estimer le comportement de réponse des ménages non-répondants, afin de réallouer leur poids aux ménages répondants ayant des caractéristiques proches. Une régression logistique permettant d'estimer la probabilité de répondre est réalisée afin d'estimer ce comportement de réponse en se basant sur un certain nombre de variables disponibles dans la base cartographique. Des groupes homogènes de réponse sont ensuite constitués dans le but de rendre le modèle logistique plus robuste. Au sein de chaque groupe, le poids des ménages non-répondants est redistribué aux ménages répondants appartenant au même groupe.

Un premier calage est ensuite réalisé au niveau logement, en raison du fait que la base de sondage soit une base d'adresses, et permet une première fois d'ajuster les poids sur le nombre de résidences principales.

La deuxième étape des redressements est le traitement de la non-réponse au niveau individuel. De manière analogue au traitement de la non-réponse au niveau logement, une régression logistique a été réalisée à partir de variables individuelles, en prenant en compte dans la pondération la correction réalisée au niveau logement.

Enfin, la troisième étape des redressements consiste en un calage sur marges au niveau individuel, à partir des données issues du recensement de la population. Un calage sur marges externes par la méthode Logit bornée a été utilisé en mobilisant quatre marges :

- la population par sexe et tranche d'âge ;
- la population par niveau de diplôme le plus élevé selon les regroupements de la nomenclature ISCED (0-2, 3-4, 5-8) ;
- le lieu de naissance agrégé (à Mayotte, en France à l'exclusion de Mayotte, à l'étranger) ;
- le nombre de résidences principales (une nouvelle fois en raison du tirage réalisé à partir de la base cartographique).

2 Estimation de la précision

2.1 Méthodologie d'estimation

Produire un intervalle de confiance pour les indicateurs des différents modules de FLV revient à s'intéresser à la variance de ces indicateurs, c'est-à-dire à la variance imputable à chaque degré du plan de sondage menant à l'échantillon de logements enquêtés. De plus, à ces différents degrés doivent être ajoutées la variance liée au mécanisme de non-réponse des logements puis la variance liée à la non-réponse des individus finalement interrogés, menant à la base de données sur laquelle sont produits les indicateurs.

Ainsi, le mécanisme de constitution de l'échantillon d'étude peut être résumé en un ensemble de degrés, où chaque étape correspond au final à un degré de sondage distinct :

- tirage des adresses ;
- tirage des logements dans les grandes adresses ;
- tirage des logements répondants ;
- tirage d'un individu dans le logement ;
- tirage des individus répondants ;
- éventuellement, tirage des formations.

La méthodologie adoptée afin de calculer la variance liée à l'ensemble du plan de sondage est celle de la formule de Rao, qui permet d'obtenir des expressions analytiques des variances dans le cadre de plans de sondage à plusieurs degrés. Ici, la formule de Rao est utilisée à chaque niveau du plan de sondage de manière itérative. À chaque degré, la formule est employée en considérant les degrés précédents comme un tout.

La méthode retenue ici est identique à celle mobilisée à l'égard des autres données issues de l'enquête FLV, à savoir le territoire métropolitain et les autres départements d'Outre-mer, les différences étant justifiées par des bases de données et des plans de sondage différents.

2.2 Formule de Rao

Pour rappel, la formule de Rao s'énonce comme suit. Soit p un plan de sondage à deux degrés. On échantillonne au premier degré S_1 de taille N_{up} , selon un plan de sondage P_1 avec $\pi^{(1)}$ la probabilité d'inclusion d'une unité primaire up . Au deuxième degré, on sélectionne indépendamment dans chaque unité primaire un échantillon S_2 d'unités secondaires selon un plan de sondage P_2 , avec $\pi^{(2/1)}$ la probabilité d'inclusion d'unité secondaire us , conditionnellement au tirage du premier degré.

On peut définir \hat{Y} l'estimateur d'Horvitz-Thompson du total Y de la variable y , tel que :

$$\hat{Y} = \sum_{up \in S_1} \frac{\hat{y}(up)}{\pi_{up}^{(1)}} \text{ avec } \hat{y}(up) = \sum_{k \in up \cap S_2} \frac{y_k}{\pi_k^{(2/1)}}$$

Sachant l'indépendance entre les tirages réalisés au sein des unités primaires, et en supposant que :

- il existe un estimateur $\hat{V}_1(\hat{Y})$ sans biais de la variance au premier degré d'échantillonnage, pouvant s'écrire sous une forme quadratique

$$Q(y_{(1)}, y_{(2)}, \dots, y_{(N_{up})}) = \sum_{i \in 1 \dots N_{up}} \sum_{j \neq i} q_{ij} y_{(i)} y_{(j)} + \sum_i q_i y_{(i)}^2 ;$$

- pour toute unité primaire $up \in S_1$, il existe :
 - un estimateur sans biais $\hat{y}(up)$ du total sous le deuxième degré $y(up) = \sum_{k \in up} y_k$;
 - un estimateur sans biais de la variance de l'estimateur $\hat{y}(up) : \hat{V}_{2/1}(\hat{y}(up))$.

Alors sous ces hypothèses, en utilisant la formule des variances conditionnelles, l'estimateur de la variance de l'estimateur de Horvitz-Thompson \hat{Y} sous le plan à deux degrés peut être décomposé en isolant l'effet de chaque degré du plan de sondage.

Cette variance est alors estimée sans biais par :

$$\hat{V}(\hat{Y}) = Q(\hat{y}_{(1)}, \hat{y}_{(2)}, \dots, \hat{y}_{(N_{up})}) + \sum_{up \in S_1} \left(\frac{1}{(\pi_{up}^{(1)})^2} - q_{up} \right) \hat{V}_{2/1}(\hat{y}(up))$$

2.3 Manipulation du package *gustave*

Dans le cadre de l'estimation de variance, le package *gustave* développé en R par Martin Chevalier en 2018 et maintenu par l'Insee offre toute une gamme d'outils allant de l'implémentation des estimateurs de variance à la production simple des informations nécessaires aux calculs de précision, dont les intervalles de confiances. Cet outil a donc été employé pour estimer la variance de l'ensemble des indicateurs issus de l'enquête FLV à Mayotte.

2.3.1 Formule de variance

La formule globale de variance des estimateurs de l'enquête FLV à Mayotte peut être résumée comme l'imbrication de cinq ou six formules de variance selon les cas, par l'intermédiaire de la formule de Rao.

La variance de premier degré, liée au tirage des adresses, peut être estimée par l'intermédiaire de la formule de Deville et Tillé pour les tirages équilibrés, sans stratification. En réalité, un tri par une variable géographique faisant intervenir les îlots a été réalisé en préalable du tirage de la première vague, entraînant de ce fait une stratification implicite. Lors du tirage de la deuxième vague de collecte, une stratification par secteur d'activité des enquêteurs (SAE) a été utilisée. Afin d'être le plus conservateur possible et en l'absence de lien suffisamment fort entre îlots et SAE, il est ainsi considéré qu'aucune stratification n'a été réalisée.

Au deuxième degré, l'échantillon des logements situés à une grande adresse est produit à partir d'un plan de sondage de taille fixe, et la variance de ce degré de tirage peut être estimée par l'estimateur de Sen-Yates-Grundy.

Le tirage des logements répondants au troisième degré peut être approché par un plan de sondage poissonien, dont la variance est estimée par la formule de Horvitz-Thompson.

Au quatrième et degré, dans chacun des logements répondants, un individu est tiré de manière aléatoire, afin de répondre au questionnaire. Ce tirage est pris en compte par l'intermédiaire d'un plan de sondage poissonien dont la variance est estimée par l'estimateur de Horvitz-Thompson.

De même, le sondage des individus interrogés répondants est approché par un plan poissonien, dont la variance est estimée par la formule de Horvitz-Thompson.

En pratique, l'implémentation sous R du calcul de la variance est facilitée par l'utilisation des outils proposés par le package *gustave*. Sous réserve que les variances y aient déjà été implémentées, il est aisé de pré-calculer la variance relative à chaque degré du plan de sondage, en fournissant les arguments adaptés (**exemple 1**). Une force du package est qu'il permet de réaliser cette implémentation sans pour autant définir directement l'indicateur pour lequel devra être estimée la précision. Les différents objets pré-calculés peuvent ensuite être associés dans un objet **precalcul_tous_degres** d'une manière similaire à ce qui est fait de façon analytique via la formule de Rao.

Exemple 1² : Implémentation avec *gustave* du calcul de la variance aux différents degrés du plan de sondage FLV à Mayotte

```
# PREMIER DEGRÉ : tirage équilibré des adresses
precalcul_degre_1 <- varDT(y = NULL,
                          x = donnees_equilibrage,
                          pik = liens_degres$pi_tirage_adresses,
                          id = liens_degres$id_adresses)

# DEUXIÈME DEGRÉ : tirage des logements dans les grandes adresses
precalcul_degre_2 <- var_srs(y = NULL,
                             pik = liens_degres$pi_tirage_logements,
                             id = liens_degres$id_logements)

# TROISIÈME DEGRÉ : tirage des logements répondants
precalcul_degre_3 <- var_pois(y = NULL,
                              pik = liens_degres$pi_tirage_log_rep,
                              id = liens_degres$id_logements)

# QUATRIÈME DEGRÉ : tirage des individus interrogés
precalcul_degre_4 <- var_pois(y = NULL,
                              pik = liens_degres$pi_tirage_ind_interrog,
                              id = liens_degres$id_individus)

# CINQUIÈME DEGRÉ : tirage des individus répondants
precalcul_degre_5 <- var_pois(y = NULL,
                              pik = liens_degres$pi_tirage_ind_rep,
                              id = liens_degres$id_individus)
```

2.3.2 Tirage des formations détaillées

Au cours de la passation du questionnaire, plusieurs formations suivies par la personne interrogée font l'objet de questions complémentaires. Cinq formations sont ainsi détaillées et deux d'entre elles font l'objet de questions supplémentaires. La prise en compte de la variance liée à ces indicateurs s'apparente à un degré de sondage supplémentaire. Son implémentation sous R peut être réalisée à peu de coûts, en se basant sur les éléments similaires à ceux fournis en exemple.

2 Plusieurs éléments de l'exemple ont été simplifiés pour en faciliter la compréhension.

2.3.3 Calage, données techniques et *wrapper* de variance

Il est nécessaire de prendre en compte l'effet du calage sur marges, permettant *in fine* d'assurer la cohérence avec d'autres résultats et d'améliorer la précision des estimateurs. La prise en compte du calage dans l'estimation de la variance s'effectue de manière classique, en considérant l'estimateur de la variance du total des résidus de la régression des indicateurs sur les variables de calage.

Le package *gustave* propose également d'implémenter une fonction de calcul des résidus de cette régression, sans nécessairement tout définir à priori. Ainsi sont stockés dans l'objet **precalcul_residus_calage** les éléments nécessaires au calcul (poids, variables de calage) en attente de l'indicateur à prendre en compte (**exemple 2**). La fonction d'estimation de la variance totale **estim_variance** peut ensuite être définie, reprenant d'une part l'objet **precalcul_residus_calage**, d'autre part les informations relatives aux différents degrés de sondage stockés dans l'objet **precalcul_tous_degrés**.

Enfin, le *wrapper* de variance **estimation_variance_FLV_Mayotte** peut être créé. Il comprend tout un ensemble d'informations allant des données d'enquête à la fonction d'estimation de la variance en passant par les poids des individus calés.

Exemple 2³ : Implémentation avec *gustave* du calcul des résidus de calage, de l'estimateur global de la variance et du wrapper de variance dans FLV à Mayotte

```
#Prise en compte du calage
precalcul_residus_calage <- res_cal(y = NULL,
                                   x = variables_calage,
                                   w = poids_pre_calage,
                                   id = liens_degrés$id_individus)

# Fonction d'estimation de la variance d'un estimateur calé (plan FLV DOM hors Mayotte)
estim_variance <- fonction(y, precalcul_tous_degrés, liens_degrés, precalcul_residus_calage){
  #Calcul des résidus
  res_y <- res_cal(y = y,
                  precalc = precalcul_residus_calage)

  #Calcul de la variance sur les résidus
  variance_totale <- var_Rao_multiple(y = res_y,
                                     precalc = precalcul_tous_degrés,
                                     links = liens_degrés,
                                     id = liens_degrés$id_individus)

  return(variance_totale)
}

# Préparation des données pour le wrapper de variance
# On inclut dans l'objet 'donnees_techniques' :
# - les liens entre les différents degrés 'liens_degrés'
# - les pre-calculs de variance 'precalcul_tous_degrés'
# - le prec-calcul des résidus de calage 'precalcul_residus_calage'

# Définition du wrapper de variance
```

3 Plusieurs éléments de l'exemple ont été simplifiés pour en faciliter la compréhension.

```
estimation_variance_FLV_Mayotte <- define_variance_wrapper(
  variance_function = estim_variance,
  reference_id = donnees_techniques$id_individu,
  reference_weight = poids_individus_cales,
  default_id = "id",
  technical_data = donnees_techniques)
```

3 Résultats d'estimation et enseignements

Une fois le *wrapper estimation_variance_FLV_Mayotte* défini, il peut facilement être utilisé de multiples façons. Il s'agit d'une fonction prête à l'emploi, pouvant être appliquée à n'importe quelle variable de l'enquête sur des champs plus ou moins restreint (**exemple 3**).

Exemple 3 : Estimations avec *gustave* d'indicateurs de l'enquête FLV à Mayotte

```
#Taux de participation à l'éducation formelle et à l'apprentissage (FED) – 18-24 ans
estimation_variance_FLV_Mayotte(data = donnees_enquete_mayotte,
  mean(indicatrice_FED),
  where = between(AGE, 18, 24))

#Taux de participation à l'éducation formelle et à l'apprentissage (FED) – 18-24 ans, femmes
estimation_variance_FLV_Mayotte(data = donnees_enquete_mayotte,
  mean(indicatrice_FED),
  where = between(AGE, 18, 24) & (SEX == "femmes"))
```

Le tableau suivant présente l'ensemble des indicateurs relatifs à l'enquête FLV à Mayotte⁴ dont la précision a été calculée (tableau).

Tableau : Estimation des indicateurs de formation FLV à Mayotte

Indicateur	Estimation	Ecart-type	Coeff. de variation	IC (95 %)	Nb rép.	
Taux de participation à l'éducation formelle et à l'apprentissage	18-24 ans	24,20	13,60	0,03	[0,18 ; 0,31]	37
	18-24 ans, femmes	18,60	24,35	0,05	[0,1 ; 0,27]	21
	18-24 ans, hommes	31,40	15,41	0,05	[0,22 ; 0,41]	16
	25-69 ans	2,00	39,76	0,01	[0 ; 0,03]	17
Taux de participation à l'éducation non formelle et à l'apprentissage	25-69 ans	13,00	14,88	0,02	[0,09 ; 0,17]	109
	25-69 ans, femmes	13,40	21,54	0,03	[0,08 ; 0,19]	73
	25-69 ans, hommes	12,50	19,08	0,02	[0,08 ; 0,17]	36
	18-24 ans	18,60	18,49	0,03	[0,12 ; 0,25]	29
	25-34 ans	14,20	21,58	0,03	[0,08 ; 0,2]	39
	35-54 ans	13,90	20,80	0,03	[0,08 ; 0,2]	60
	55-69 ans	6,30	66,16	0,04	[-0,02 ; 0,14]	10

⁴ Le package *gustave* a également été utilisé pour estimer la précision dans les autres DOM ainsi qu'en France métropolitaine.

	25-69 ans, faible niveau d'éducation (ISCED 0-2)	7,20	22,83	0,02	[0,04 ; 0,1]	44
	25-69 ans, niveau moyen d'éducation (ISCED 3-4)	25,20	26,41	0,07	[0,12 ; 0,38]	36
	25-69 ans, niveau élevé d'éducation (ISCED 5-8)	25,20	26,87	0,07	[0,12 ; 0,38]	29
	25-69 ans, en emploi	22,50	15,79	0,04	[0,16 ; 0,3]	76
	25-69 ans, sans emploi	8,60	41,61	0,04	[0,02 ; 0,16]	25
	25-69 ans, en dehors de la population active	3,30	58,29	0,02	[0 ; 0,07]	8
Part des activités liées à l'emploi parmi les activités d'apprentissage non formel	25-69 ans	85,30	6,50	0,06	[0,74 ; 0,96]	96
	18-24 ans	94,90	4,49	0,04	[0,87 ; 1,03]	26
Taux de participation à l'apprentissage informel, age 18-69 - %		35,10	6,05	0,02	[0,31 ; 0,39]	327
Montant moyen payé par un participant pour toutes les activités d'apprentissage non formel, age 18-69 - EUR		705,03	47,25	333,12	[52,12 ; 1357,94]	20
Nombre d'heures d'enseignement consacrées par un participant à l'activité d'éducation formelle la plus récente, age 18-69 - heures		584,31	20,17	117,83	[353,36 ; 815,26]	33
Nombre d'heures d'enseignement consacrées par un participant aux activités d'apprentissage non formel, age 18-69 - heures		197,76	23,69	46,85	[105,93 ; 289,59]	116

Lecture : Le taux de participation à l'éducation formelle est estimé à 24,2 % chez les individus âgés de 18 à 24 ans résidant à Mayotte.

Champ : Individus et logements du champ FLV.

Source : Spécifications FLV et calculs PIMEN.

Bibliographie

Deville, JC et Särndal, Carl-Erik (1992), *Calibration Estimators in Survey Sampling*, *Journal of the American Statistical Association* 87 (418): 376-82.

Deville, J-C. et Tillé, Y. (2004), *Efficient Balanced Sampling : The Cube Method*, *Biometrika*, Vol 91, No 4, pp. 893-912.

Deville, JC. et Tillé, Y. (2005). *Variance approximation under balanced sampling*, *Journal of Statistical Planning and Inference*, Vol 128, pp 569-591.

RAO, J.N.K. (1975). *Unbiased variance estimation for multi-stage designs*. *Sankhyā, série C*, 37, 133-139.

Fabre, E. (2022), *Note de tirage de l'enquête FLV dans les DOM*, Note interne Insee, note n°2022_11988_DR974-SES.

Garoche, B. (2023), *Note de tirage de l'échantillon de la deuxième vague collectée en 2023 de l'enquête FLV Mayotte 2022-2023*, Note interne Insee, note n°2022_11988_DR974-SES.

Rivière, J. (2023), *Travaux de redressement de l'enquête FLV 2022-2023*, Note interne Insee, note n°2022_11988_DR974-SES.

Larbi, K. (2024), *Estimation de variance dans l'enquête Formation tout au long de la vie 2023*, Note interne Insee, note n°2024_11880_DG75_L110.

Loiseau, R. (2024), *Précision des estimations de l'enquête FLV dans les DOM*, Note interne Insee, à paraître.