Traitement des valeurs influentes dans les enquêtes

Cyril Favre-Martinoz & Thomas Deroyon

***Résumé* — L'objectif de cette note méthodologique est de décrire le traitement des valeurs influentes dans les enquêtes. Ce document est découpé en cinq grandes parties. Dans une première partie, nous détaillons le cadre théorique et mettons en évidence les différences entre valeurs aberrantes et valeurs influentes. Dans une deuxième partie, nous présentons les causes de la présence des valeurs influentes dans les enquêtes et nous donnons quelques bonnes pratiques à adopter au moment de la création de l'enquête afin de limiter le problème des valeurs influentes. La troisième partie est consacrée à la présentation d'un outil permettant de mesurer l'influence : le biais conditionnel. Dans une quatrième partie, nous détaillons les méthodes permettant de traiter le problème de valeurs influentes. Enfin, dans une dernière partie nous donnons des exemples de mise en œuvre de ces méthodes à l'INSEE.**

I. Cadre théorique et contextes d'application

A. Contexte

Dans les enquêtes auprès des entreprises, il est courant de collecter des variables économiques dont la distribution est fortement asymétrique. Par exemple, dans le cas du chiffre d'affaires, on observe une variable d'intérêt positive, dont la distribution est étalée vers la droite. Dans ce contexte, on est souvent confronté à la présence de valeurs influentes dans l'échantillon tiré. Ces dernières sont habituellement de très grandes valeurs dont la présence dans l'échantillon tend à rendre les estimateurs classiques (par exemple, l'estimateur par dilatation) très instables. Les estimateurs robustes sont construits de manière à limiter l'impact des valeurs influentes, ce qui conduit à des estimateurs plus stables mais potentiellement biaisés. L'objectif de l'estimation robuste consiste à détecter les valeurs influentes, puis à développer des procédures d'estimation robustes dont l'erreur quadratique moyenne est significativement inférieure à celle des estimateurs classiques en présence de valeurs influentes. De plus, on souhaiterait que ces estimateurs ne souffrent pas d'une perte d'efficacité importante en l'absence de valeurs influentes. Le traitement des valeurs influentes permet donc habituellement d'obtenir un compromis entre le biais et la variance.

B. La distinction entre valeurs aberrantes et valeurs influentes

Il est important de distinguer dans notre échantillon, deux types d'unités aberrantes : les unités aberrantes représentatives et les unités aberrantes non représentatives. Ce concept d'unités représentatives a été introduit et discuté par Chambers (1986). Les unités représentatives, qui seront dans la suite considérées comme potentiellement influentes, sont des unités dont la valeur collectée sur l'échantillon est correcte et n'est pas considérée comme unique au sens où il est probable qu'il existe dans notre population U d'autres unités ayant une valeur collectée du même ordre de grandeur. Dans le cas de l'estimation d'un paramètre de population finie comme un total, ces unités ont une importance considérable dans l'estimation de celui-ci et on ne peut pas se permettre de leur mettre un poids égal à 1, car cela reviendrait à les considérer comme uniques. Les valeurs aberrantes non représentatives sont des unités dont la valeur collectée est erronée, à cause d'un dysfonctionnement dans le processus de collecte : un cas classique est le chiffre d'affaires d'une entreprise indiqué en euros au lieu d'être indiqué en milliers d'euros. Le traitement de ce type d'unité peut être réalisé à l'étape d'apurement des données, notamment par des processus d'imputation : on peut imputer le chiffre d'affaires que l'on estime erroné par le chiffre d'affaires obtenu lors d'une précédente enquête. Ces unités aberrantes sont de fait uniques, et on peut leur attribuer un poids de 1 dans la suite du processus d'estimation ou corriger leur valeur si on est capable d'identifier l'erreur.

C. Définition des notions de configuration et de valeur influente

Avant de définir la notion de valeur influente de façon générale, on introduit le concept de configuration :

Une configuration $\mathscr{C}$ est définie par le quadruplet suivant :
(1) une variable d'intérêt y ;
(2) un paramètre d'intérêt ;
(3) un plan de sondage ;
(4) un estimateur.

Le concept de configuration est une notion centrale dans la mesure où une unité est influente dans une configuration donnée ; c'est-à-dire qu'une unité est influente pour un plan, un paramètre et un estimateur donnés. Dans une configuration $\mathscr{C}$ donnée, une valeur sera définie comme **influente** si elle a un impact significatif sur l'erreur quadratique moyenne de l'estimateur considéré. Un exemple illustrant les notions de configuration, d'unité influente et de biais conditionnel est donné dans la section (3.C).

II. Pourquoi observe-t-on des valeurs influentes dans nos enquêtes et comment se prémunir contre leur apparition ?

A. Les variables d'intérêt à distributions asymétriques

Dans les enquêtes auprès des entreprises, il est courant d'observer des variables d'intérêt comme le chiffre d'affaires dont la distribution est fortement asymétrique. Ainsi certaines entreprises ont une contribution très importante à l'agrégat que l'on souhaite mesurer. Le fait de sélectionner, ou non, une de ces ≪ grosses ≫ unités a un impact important sur l'estimateur.

B. Les migrants inter-strates ou "Strata jumpers"

Un deuxième problème conduisant à la présence de valeurs influentes dans l'échantillon est celui des "strata jumpers"

©Insee

qui survient lorsque l'information de stratification recueillie sur le terrain est différente de celle disponible sur la base de sondage. Ces différences sont habituellement dues à des imperfections dans la base de sondage (par exemple, dans le cas d'une base un peu datée). Un stratum jumper est une unité qui n'appartient pas à la strate à laquelle elle aurait dû appartenir si l'information sur la base de sondage était correcte. Si une unité avec une grande valeur est assignée à une strate non-exhaustive, elle combinera alors une grande valeur de la variable d'intérêt et éventuellement un grand poids de sondage, ce qui la rendrait potentiellement très influente. En pratique, il n'est pas rare d'observer entre 5% et 10% de "stratum jumpers". Ce pourcentage est d'autant plus important que la base de sondage est ancienne.

C. Une mauvaise corrélation entre les poids de sondage et la variable d'intérêt

Il est possible de se prémunir contre l'impact des valeurs influentes à l'étape du plan de sondage en sélectionnant d'office les unités potentiellement influentes. Par exemple, dans les enquêtes auprès des entreprises, il est de coutume d'utiliser un plan stratifié aléatoire simple sans remise comportant une ou plusieurs strates exhaustives composées habituellement des grandes unités. Malheureusement, il est rarement possible d'éliminer complètement le problème des valeurs influentes à l'étape du plan de sondage. En effet, les strates dans les enquêtes auprès des entreprises sont habituellement formées d'une variable de taille (par exemple, l'effectif salarié au 31 décembre de l'année précédente) et d'une variable de classification de l'activité (par exemple, le code APE). Dans une enquête recueillant des dizaines de variables d'intérêt, il n'est pas improbable que certaines d'entre elles soient peu ou pas liées aux variables de stratification, pouvant alors conduire à la présence de valeurs influentes.

Afin de se prémunir contre le problème des valeurs influentes, il est également important de contrôler les facteurs de correction des poids issus des méthodes de redressement :
- A l'étape de la correction de la non-réponse, l'utilisation de classes de repondération permet de se prémunir contre des variations extrêmes des facteurs de correction.
- A l'étape de calage, il est possible de borner le facteur de variation de poids via l'utilisation d'une fonction de distance adaptée.

III. Comment quantifier l'influence d'une unité ?

A. Le biais conditionnel : un outil pour mesurer l'influence

La notion de valeur influente est relativement vague et il est nécessaire de disposer d'un outil permettant de mesurer l'influence tout en tenant compte du plan de sondage. Dans le cas d'une approche sous le plan, la notion de biais conditionnel a été développé dans deux articles de Moreno-Rebollo et al. (1995, 1999). Cette notion de biais conditionnel a été reprise par Beaumont et al. (2013) afin de quantifier l'influence sous le plan d'une unité afin de construire ensuite des estimateurs robustes. Soit $U=(1,\dots,k,\dots,N)$ une population finie, $P(.)$ un plan de sondage défini sur U et Y la variable d'intérêt à observer sur la population. Soit θ le paramètre d'intérêt et $\hat{\theta}$ un estimateur de θ. Le biais conditionnel d'une unité échantillonnée i associé à l'estimateur $\hat{\theta}$ est défini par : $B_i^{\hat{\theta}}(I_i=1)=\mathbb{E}_P\left(\hat{\theta}|I_i=1\right)-\mathbb{E}_P\left(\hat{\theta}\right)$, où I_i est la variable indicatrice d'appartenance à l'échantillon qui prend la valeur 1 si l'unité i est dans l'échantillon et 0 sinon. De façon similaire, on définit le biais conditionnel d'une unité i non échantillonnée associé à l'estimateur $\hat{\theta}$ par :

$$B_i^{\hat{\theta}}(I_i=0)=\mathbb{E}_P\left(\hat{\theta}|I_i=0\right)-\mathbb{E}_P\left(\hat{\theta}\right).$$

Le biais conditionnel est une mesure d'une influence car il permet d'observer l'impact moyen engendré sur l'estimateur suivant que l'unité i appartienne ou non à l'échantillon. Il est important de noter que le biais conditionnel d'une unité non échantillonée est inconnu, et impossible à estimer car il fait intervenir les valeurs de la variable d'intérêt hors de l'échantillon. On ne peut donc pas se prémunir de l'influence des unités non échantillonnées.
On va calculer explicitement l'influence sur l'estimateur de Horvitz-Thompson défini par :

$$\hat{t}_{y\pi}=\sum_{j\in S}\frac{y_j}{\pi_j}.$$

où π_j est la probabilité d'inclusion de l'unité j. On désigne par $d_j=\frac{1}{\pi_j}$ le poids de sondage de l'unité j. Le biais conditionnel d'une unité échantillonnée i pour l'estimateur de Horvitz-Thompson est donné par :

$$\begin{aligned} B_i^{HT}(I_i=1) &= E_p\left(\hat{t}_{y\pi}|I_i=1\right)-t_y \\ &= \sum_{j\in U}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)y_j. \end{aligned} \qquad (1)$$

Etant donné que l'expression du biais conditionnel fait intervenir les probabilités d'inclusion d'ordre 1 et 2, il tient compte du plan de sondage.

B. Les propriétés du biais conditionnel

1) On constate qu'une unité ayant une probabilité d'inclusion égale à 1 a un biais conditionnel nul, i.e, l'influence de cette unité est égale à zéro. Ainsi on comprend ici que l'utilisation d'une strate exhaustive prend tout son sens.

2) Il est également important de noter que l'erreur d'échantillonnage de l'estimateur de Horvitz-Thompson $\hat{t}_{y\pi}-t_y$ peut se décomposer de la façon suivante

$$\hat{t}_{y\pi}-t_y=\sum_{i\in S}B_i^{HT}(I_i=1)+\sum_{i\in U\setminus S}B_i^{HT}(I_i=0) \qquad (2)$$

si

$$\sum_{i\in U}(I_i-\pi_i)\,a_i=0, \qquad (3)$$

où $a_i=(1-\pi_i)^{-1}\left\{B_i^{HT}(I_i=1)-(d_i-1)\,y_i\right\}$.
On peut montrer que la condition (3) est vérifiée pour un plan de sondage poissonien. La décomposition (2) est approximativement respectée pour un plan de sondage stratifié aléatoire sans remise ou un plan de sondage à grande entropie de taille fixe. Dans le cas où la décomposition (2) est valable, le biais conditionnel peut être vu comme la contibution de l'unité i à l'erreur d'échantillonnage $\hat{t}_{y\pi}-t_y$.

3) On a également la propriété suivante pour n'importe quel plan de sondage $P(.)$:

$$Var_p(\hat{t}_{y\pi})=\sum_{j\in U}\sum_{k\in U}\frac{y_jy_k}{\pi_j\pi_k}\Delta_{jk}=\sum_{i\in U}B_i^{HT}(I_i=1)y_i. \qquad (4)$$

©Insee

où Δ_{jk} est la matrice de variance-covariance des indicatrices I_j et I_k.

La variance de l'estimateur de Horvitz-Thompson est donc directement reliée au biais conditionnel et on constate qu'une unité ayant un fort biais conditionnel contribuera fortement à la variance. De plus, elle contribuera d'autant plus fort à la variance que la valeur de la variable d'intérêt y_i sera élevée.

C. *Exemple pour deux plans de sondage particuliers*

Considérons une population de taille 5000 pour laquelle on observe les chiffres d'affaires fictifs en milliers d'euros y, rangés par ordre croissant :

$$y_1 = 0, y_2 = 500, y_3 = ... = y_{4999} = 500 \text{ et } y_{5000} = 2000$$

Dans ce cas, la moyenne dans la population $\bar{y}_U$ est égale à 500.2. Supposons que l'on se trouve dans une des deux configurations suivantes :

$\mathscr{C}_1$: (Chiffre d'affaires, Chiffre d'affaires total, sondage aléatoire simple sans remise, estimateur de Horvitz-Thompson)

$\mathscr{C}_2$: (Chiffre d'affaires, Chiffre d'affaires total, Tirage poissonien avec probabilités égales $\pi_k = \frac{n}{N}, k \in U$, estimateur de Horvitz-Thompson)

Afin de faire le lien entre le biais conditionnel et l'instabilité des estimateurs, nous rappelons dans le tableau 1, le biais conditionnel associé à une unité sélectionnée et les formules de variance pour l'estimateur de Horvitz-Thompson.

	Variance	Biais conditionnel de l'unité i
Sondage aléatoire simple sans remise	$N^2 \frac{(1-\frac{n}{N})}{n} S^2_{y_U}$	$\frac{N}{N-1}(\frac{N}{n}-1)(y_i - \bar{y}_U)$
Tirage Poissonien	$\sum_{k \in U} \frac{(1-\pi_k)y_k^2}{\pi_k}$	$(d_i - 1)y_i$

Tableau 1 : Résumé des formules de variance et du biais conditionnel pour l'estimateur de Horvitz-Thompson

Dans le cas d'un sondage aléatoire simple sans remise, la première unité dont le chiffre d'affaires est égale à 0 contribue fortement à la variance de l'estimateur de Horvitz-Thompson si elle est sélectionnée (elle contribue à une valeur élevée de la dispersion $S^2_{y_U}$, alors que dans le cas poissonien, la première unité ne contribue pas à la variance de l'estimateur de Horvitz-Thompson (le terme $k = 1$ est nul dans la formule de variance associée au Tirage Poissonien). Ainsi, l'influence d'une unité dépend fortement du plan utilisé. On peut le voir directement pour chaque unité à l'aide du biais conditionnel : dans le premier cas, le biais conditionnel est très élevé puisque la valeur 0 est très éloignée de la moyenne $\bar{y}_U = 500,2$. Alors que dans le cas du tirage poissonien à probabilités égales $\pi_k = \frac{n}{N}, k \in U$, le biais conditionnel est nul, car $y_1 = 0$ et donc l'influence de la première unité est nulle dans la deuxième configuration, alors qu'elle est élevée dans la première configuration. Enfin, l'unité ayant pour valeur $y_{5000} = 2000$, est influente pour les deux plans de sondage.

IV. Comment traiter le problème des valeurs influentes ?

A. *La méthode classique de Winsorisation*

En pratique, une méthode particulièrement utilisée est la winsorisation, qui consiste à réduire à un certain seuil les valeurs trop élevées dans l'échantillon. Dans le cas de la winsorisation, une unité est considérée comme un influence si le produit de son poids par sa valeur dépasse un certain seuil. Dans la littérature, on distingue deux types de winsorisation. La winsorisation standard, ou encore appelée winsorisation de type I dans le cas d'un sondage aléatoire simple sans remise, consiste à réduire la valeur des unités dépassant un certain seuil en tenant compte de leur poids. Soit $\tilde{y}_i$ la valeur de la variable y pour l'unité i après winsorisation. On a

$$\tilde{y}_i = \begin{cases} y_i & \text{si } d_i y_i \leq K \\ \frac{K}{d_i} & \text{si } d_i y_i > K \end{cases} \quad (5)$$

où $K > 0$ est le seuil de winsorisation. L'estimateur winsorisé standard du total t_y est donné par

$$\hat{t}_s = \sum_{i \in S} d_i \tilde{y}_i \quad (6)$$

Une écriture alternative consiste à exprimer $\hat{t}_s$ comme une somme pondérée des valeurs initiales au moyen de poids modifiés :

$$\hat{t}_s = \sum_{i \in S} \tilde{d}_i y_i,$$

où

$$\tilde{d}_i = d_i \frac{min\left(y_i, \frac{K}{d_i}\right)}{y_i}. \quad (7)$$

Si $min\left(y_i, \frac{K}{d_i}\right) = y_i$ (c'est-à-dire que l'unité i n'est pas influente), alors $\tilde{d}_i = d_i$. Le poids d'une unité non influente n'est donc pas modifié. Par contre, le poids modifié d'une unité influente est inférieur à d_i et peut même être inférieur à 1. Il convient de noter qu'une unité affichant une valeur $y_i = 0$ ne pose pas de problème particulier puisque sa contribution au total estimé, $\hat{t}_s$, est nulle. Dans ce cas, on peut assigner une valeur arbitraire au poids modifié $\tilde{d}_i$. D'un point de vue pratique, il est peu commode d'attribuer un poids inférieur à 1 à une unité, car on souhaite qu'au minium elle se représente. C'est pourquoi, on préfère en général la winsorisation de Dalén-Tambay, appelée winsorisation de type 2 dans le cas d'un sondage aléatoire simple sans remise. On définit les valeurs de la variable d'intérêt après winsorisation par

$$\tilde{y}_i = \begin{cases} y_i & \text{si } d_i y_i \leq K \\ \frac{K}{d_i} + \frac{1}{d_i}(y_i - \frac{K}{d_i}) & \text{si } d_i y_i > K \end{cases} \quad (8)$$

Cela conduit à l'estimateur winsorisé du total t_y :

$$\hat{t}_{DT} = \sum_{i \in S} d_i \tilde{y}_i. \quad (9)$$

Comme pour $\hat{t}_s$, une écriture alternative consiste à exprimer $\hat{t}_{DT}$ comme une somme pondérée des valeurs initiales au moyen de poids modifiés :

$$\hat{t}_{DT} = \sum_{i \in S} \tilde{d}_i y_i,$$

où

$$\tilde{d}_i = 1 + (d_i - 1) \frac{min\left(y_i, \frac{K}{d_i}\right)}{y_i}. \quad (10)$$

Comme pour l'estimateur winsorisé standard, le poids d'une unité non-influente n'est pas modifié. Encore une fois, une unité affichant une valeur $y_i = 0$ ne pose pas de problème particulier puisque sa contribution au total estimé, $\hat{t}_{DT}$, est

©Insee

nulle. Dans ce cas, on peut assigner une valeur arbitraire au poids modifié $\tilde{d}_i$.

B. Le choix du seuil : un choix déterminant dans le compromis biais-variance

La quasi-intégralité des méthodes robustes font intervenir un seuil. Le choix du seuil K est très important, car il permet de réaliser le compromis biais-variance pour l'estimateur robuste. Il existe en pratique trois façons de choisir ce seuil :
- à partir d'un dire d'expert : ce choix est extrêment risqué car il peut engendrer un estimateur robuste moins efficace en termes d'erreur quadratique que l'estimateur non robuste initial.
- en minimisant l'erreur quadratique moyenne estimée de l'estimateur robuste ; on peut citer par exemple la méthode de Kokic et Bell (1994) qui s'applique dans le cas de l'estimation du total d'une variable d'intérêt positive à partir d'un échantillon sélectionné par un sondage aléatoire simple stratifié à un degré en supposant que, dans chaque strate, les valeurs de la variable d'intérêt sont des réalisations d'une même loi de probabilité et que nous disposons d'observations de la variable d'intérêt dans chaque strate indépendantes de l'échantillon (par exemple issues de la base de sondage ou d'une enquête précédente).
- en choississant le seuil minimisant le maximum des influences calculées sur l'estimateur robuste : les détails de cette méthode sont donnés dans l'article de Beaumont et al. (2013).

Suivant les données disponibles dans la base de sondage ou dans les enquêtes précédentes et la complexité du plan de sondage, on peut avoir recours au point 2 ou 3. La minimisation de l'erreur quadratique moyenne de l'estimateur robuste est relativement complexe et sa mise en oeuvre n'est réalisable que pour des paramètres simples tels que le total ou la moyenne et pour des plans assez simples : sondage aléatoire simple stratifié ou poissonien. De plus, il est très difficile de généraliser cette méthode basée sur la minimisation de l'erreur quadratique moyenne pour tenir compte de la modélisation de la non-réponse ainsi que de la phase de calage. Les méthodes basées sur le biais conditionnel permettent de prendre en compte ces deux étapes indispensables dans le redressement d'une enquête. Pour plus de détails sur la généralisation des méthodes basées sur le biais conditionnel permettant de tenir compte de la phase de non-réponse et des redressements par calage, le lecteur pourra se référer aux articles de Favre-Martinoz et al. (2015, 2016).

V. Un exemple de traitement des valeurs influentes : le cas de l'enquête ESANE

Depuis 2008, les enquêtes du dispositif Esane (Élaboration des Statistiques Annuelles d'Entreprises) utilisent des techniques de winsorization suivant la méthode proposée par Kokic et Bell (1994). Cette méthode suppose de disposer de données, extérieures à l'enquête, sur la distribution de la variable winsorizée dans les strates de tirage. Le dispositif d'Esane permet de disposer du chiffre d'affaires fiscal de toutes les entreprises de la base de sondage, pour définir les seuils de winsorization qui sont utilisés, à partir de 2013, pour la winsorization. Une fois les seuils de la winsorisation déterminés pour la variable de chiffre d'affaires se pose alors la question du traitement des autres variables de la liasse fiscale des entreprises. Nous rappelons que la liasse fiscale d'une entreprise contient un grand nombre de variables, liées entre elles par de nombreuses relations comptables. Une entreprise peut n'être atypique que pour certaines de ces variables. Aussi, il serait possible de réaliser une winsorization séparée pour chaque variable de la liasse fiscale. Des seuils seraient calculés pour le chiffre d'affaires, la valeur ajoutée, l'excédent brut d'exploitation, l'investissement,... et les valeurs atypiques identifiées et traitées sur la base de ces seuils. Cette méthode risque cependant de rompre les relations comptables existant entre les variables d'une même liasse pour les unités winsorizées. Un choix raisonné consiste à calculer les poids winsorisés correspondant aux seuils déterminés par la méthode de Kokic et Bell (1994) et à utiliser ces poids pour les autres variables d'intérêt de l'enquête. Cet ajustement est efficace si les autres variables d'intérêt sont très corrélées au chiffre d'affaires. C'est le cas par exemple de la valeur ajoutée et de la masse salariale. Par contre, cette méthode peut poser problème quand il s'agit de détecter les valeurs influentes de variables peu corrélées au chiffre d'affaires, comme l'investissement, qui, de manière générale, est une variable complexe à traiter, présentant une grande variance et une faible cohérence temporelle.

References

[1] Beaumont, J.-F., Haziza, D. and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, 100, 555–569.

[2] Chambers, R. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, 81, 1063–1069.

[3] Dalén, J. (1987). Practical estimators of a population total which reduce the impact of large observations. R and D Report. Statistics Sweden.

[4] Deroyon, T. (2015). Traitement des valeurs atypiques d'une enquête par winsorization. Application aux enquêtes sectorielles annuelles. Acte des Journées de Méthodologie Statistique de l'Insee, 1er avril 2015, Paris.

[5] Favre-Martinoz, C., Beaumont, J.-F., Haziza, D.(2015) *Une méthode de détermination du seuil pour la winsorisation avec application à l'estimation pour des domaines*, Techniques d'enquête, 2015.

[6] Favre-Martinoz, C., Haziza, D., Beaumont, J.-F.(2016). *Robust Inference in Two-phase Sampling Designs with Application to Unit Nonresponse*. Scandinavian Journal of Statistics.

[7] Kokic, P.N. and Bell, P.A. (1994). Optimal Winsorizing cutoffs for a stratified finite population estimator. *Journal of Official Statistics*, 10, 419–435.

[8] Moreno-Rebollo, J.L., Muñoz-Reyez, A.M. and Muñoz-Pichardo, J.M. (1999). Influence diagnostics in survey sampling : conditional bias. *Biometrika*, 86, 923–928.

[9] Munoz-Pichardo, J., Munoz-Garcia, J., Moreno-Rebollo, J. and Pino-Mejias, R. (1995). A new approach to influence analysis in linear models. *Sankhyā : The Indian Journal of Statistics, Series A*, 393–409.

Insee
Mesurer pour comprendre

Département des méthodes statistiques
Version n^o 1, diffusée le 10 octobre 2017

©Insee