Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de l'intérieur DFI
Office fédéral de la statistique OFS

Typologie des ménages dans la statistique STATPOP (HHTyp)

Processus de production - description courte

Neuchâtel, 2022

Éditeur:	Office fédéral de la statistique (OFS)	**Concept de mise en page:**	Section PUB
Renseignements:	info.dem@bfs.admin.ch	**Téléchargement:**	www.experimental.bfs.admin.ch
Rédaction:	Sabrina Bravo, OFS ; Estelle Crippa, OFS ; Fabienne Helfer, OFS	**Copyright:**	OFS, Neuchâtel 2022
Contenu:	Sabrina Bravo, OFS ; Estelle Crippa, OFS		
Domaine:	00 Bases statistiques des généralités		
Langue du texte original:	Français		

La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée

Introduction

Une variable « type de ménage » a été créée pour la statistique de la population et des ménages (STATPOP). Elle assigne un type de ménage à tous les ménages privés de la population résidante permanente au domicile principal. Le type de ménage est une information importante non seulement dans le domaine des ménages mais aussi en lien avec d'autres thématiques. Jusqu'en 2000, la variable « Type de ménage » était déduite des informations relevées dans les recensements exhaustifs de la population qui avaient lieu tous les dix ans. Avec la mise en place du nouveau système de recensement à partir de 2010, cette variable est disponible dans le Relevé Structurel, une enquête annuelle réalisée sur un échantillon d'environ 300 000 personnes. Le « Type de ménage » selon le Relevé Structurel est basé sur les relations entre tous les membres du ménage et permet d'obtenir une estimation directe et fiable pour des groupes d'au moins 15 000 personnes. Dans les registres des habitants qui sont la base de la statistique de la population et des ménages, ces relations entre tous les membres du ménage ne sont pas disponibles. Grâce à ce projet sur la typologie des ménages, cette variable est maintenant disponible de manière exhaustive dans STATPOP. Elle permet de faire des analyses à un niveau plus fin que le Relevé Structurel, par exemple, regarder les évolutions des types de ménage à travers le temps dans les communes ou encore comparer les revenus de travail par type de ménage à un niveau géographique fin. Elle permet également d'être utilisée au niveau individuel comme variable explicative pour des modèles statistiques et ainsi aider à mieux comprendre d'autres problématiques concernant différentes thématiques, comme par exemple des analyses sur la pauvreté.

La variable STATPOP « Type de ménage » est créée en plusieurs étapes. Dans un premier temps, différentes sources de données sont employées pour établir les relations entre les personnes au sein de chaque ménage, puis un algorithme assigne un type de ménage en fonction des relations du ménage, pour les ménages dont toutes les relations sont connues. Pour finir, un algorithme de machine learning impute le type de ménage pour les ménages dont le type est manquant (env. 13%).

Les modalités de la variable type de ménages sont les suivantes[1] :

a. Ménage d'une personne
b. Couple marié sans enfant
c. Couple en union libre[2] sans enfant
d. Couple de même sexe[3] sans enfant
e. Parent seul avec au moins un enfant[4] de moins de 25 ans
f. Couple marié avec au moins un enfant de moins de 25 ans
g. Couple en union libre avec au moins un enfant de moins de 25 ans
h. Couple de même sexe avec au moins un enfant de moins de 25 ans
i. Autre ménage de plusieurs personnes

[1] Une deuxième typologie suivant la même procédure a été calculée en modifiant l'âge maximum considéré pour les enfants à 18 ans au lieu de 25 ans.

[2] Dans les couples en union libre sont considérés uniquement les couples de sexes opposés.

[3] Dans les couples de même sexe sont considérés les couples en union libre et les couples en partenariat enregistré.

[4] Un enfant est défini comme la fille ou le fils d'un membre du ménage.

Source des relations

Les informations des relations entre les différentes personnes d'un ménage sont reprises des sources suivantes :

a. *Registre informatisé de l'état civil* (**INFOSTAR**) : dans les données INFOSTAR se trouvent les liens « père-enfant », « mère-enfant », « épouse-époux » et « partenaires enregistrés » pour toutes les personnes qui ont eu un événement d'état civil (p. ex. être né, avoir un enfant, se marier, etc.) en Suisse, ce qui représente actuellement environ 85% de la population résidante (99,8% des Suisses et 45,5% des étrangers).

b. *Registre des diplomates et fonctionnaires internationaux* (**ORDIPRO**) : toutes les personnes ayant une autorisation de travail en Suisse délivrée par le département fédéral des affaires étrangères ainsi que les membres de leur famille et leur personnel de maison sont inscrits dans le registre ORDIPRO. Une variable donnant les relations entre la personne de référence (diplomate ou fonctionnaire international) et certains membres majeurs de sa famille, comme par exemple l'époux, l'épouse, le concubin, la concubine, les enfants, est disponible dans ce registre.

c. ***Relevé Structurel*** à partir de 2010 jusqu'à l'année n-1 : la relation de chaque « paire » des membres du ménage est relevée annuellement directement auprès d'un échantillon de la population. La relation annoncée dans le Relevé Structurel est reprise pour toutes les paires de personnes vivant toujours ensemble dans un ménage entre le moment du relevé et la production de la typologie, avec des conditions sur l'état civil pour les concubins, partenaires, époux, apparentés et non-apparentés. Tous les types de liens qui existent dans le Relevé Structurel sont ainsi repris. Le Relevé Structurel publie ses résultats après la production de la variable type de ménage dans STATPOP, c'est pourquoi la dernière année disponible est l'année n-1.

d. ***Algorithmes déterministes*** : sur la base de variables démographiques présentes dans les données de la statistique de la population et des ménages (STATPOP) entièrement basée sur les registres, il est possible de définir des relations entre deux personnes avec un résultat relativement sûr. Par exemple, deux personnes de sexe différent, mariées avec la même date de mariage, vivant dans le même ménage, seront définies comme époux-épouse. Il s'agit, par exemple, de retrouver les couples mariés qui ont immigrés et qui ne sont pas dans INFOSTAR, car leur événement d'état civil n'a pas eu lieu en Suisse.

Préparation des données

Les relations reprises des sources ci-dessus sont plausibilisées (p. ex. dans une relation parent-enfant, les deux personnes doivent avoir plus de 12 ans d'écart, un enfant ne peut avoir plus de deux parents, une personne ne peut être mariée qu'à une seule autre personne) et d'éventuelles relations symétriques sont ajoutées (dans le cas où la relation A-B est décrite, mais pas la relation B-A).

Ensuite, des **Relations éloignées** sont ajoutées. Cela consiste, à déduire, sur la base des relations plausibilisées, des relations qui ne sont pas contenues dans les registres (p.ex. grand-père, beau-frère, oncle, …). Le but est de compléter au maximum les relations, car un type de ménage ne peut être déterminé que si toutes les relations du ménage sont connues.

La contribution de chaque source pour les relations est, pour l'année 2018, la suivante : sur les environ 16 millions de « paires » de personnes au sein d'un ménage, 84,04% des relations peuvent être reprises d'une source à disposition ou calculées directement. Les relations sont ajoutées dans l'ordre ci-après : INFOSTAR (75,39%), ORDIPRO (+0,05%), Relevé Structurel (+6,12%), Algorithme déterministe (+1,78%), Relation éloignée (+0,7%).

Tableau 1 : *Relations disponibles selon la source de données*

	INFOSTAR[5]	ORDIPRO	Relevé Structurel	Algorithme déterministe	Relations éloignées
Epoux/se	X	X	X	X	
Partenaire enregistré-e	X	X	X	X	
Concubin-e		X	X	X[6]	
Père / mère	X	X	X		
Fils / Fille	X	X	X		
Frère ou demi-frère / sœur ou demi-sœur	X	X	X		
Beau-père / belle-mère (époux/se du parent)	X		X		X
Beau-fils / belle-fille (enfant de l'époux/se)	X		X		X
Grand-père / grand-mère	X	X	X		X
Petit-fils / petite-fille	X	X	X		X
Apparenté[7]		X	X		X
Non-Apparenté[8]	X		X		X

Un algorithme qui assigne un type de ménage en fonction des relations des membres du ménage est ensuite appliqué et donne un type de ménage aux ménages dont toutes les relations sont connues. En 2018, 13,78% des ménages ont un type de ménage manquant, autrement dit, pour 13,78% des ménages, au moins une de ses relations est inconnue avec les données à disposition et une imputation est nécessaire.

Imputations

Plusieurs méthodes d'imputation ont été testées, notamment un random forest sur les relations, un random forest sur les types de ménage, une procédure qui impute le type de ménage de manière déterministe, une régression multinomiale et un arbre de décision. Finalement, c'est l'arbre de décision qui a été retenu. Bien que le random forest donnait des résultats individuels légèrement meilleurs, l'arbre de décision estime mieux la qualité de ses imputations et est plus flexible pour faire des ajustements.

L'arbre de décision est créé en utilisant un ensemble d'apprentissage qui contient des ménages dont le type est connu, et des variables choisies, qui doivent être disponibles dans l'ensemble d'apprentissage, mais aussi dans l'ensemble des ménages à imputer (p. ex. nombre de personnes dans le ménage, moyenne d'âge des membres du ménage, nombre de noms de famille différents dans le ménage, taille de la commune, …). Au moyen de tests statistiques, les variables qui séparent au mieux les types de ménage sont gardées et définissent l'arbre retenu. Cet arbre est, dans un deuxième temps, appliqué aux ménages à imputer et selon les caractéristiques de ces ménages, un type est imputé.

Pour l'imputation, le Relevé Structurel joue un rôle essentiel dans la création de l'ensemble d'apprentissage. En effet, l'intersection entre l'échantillon des ménages du Relevé Structurel, tiré aléatoirement, et la sous-population des ménages à imputer de la même année peut être vue comme un domaine d'analyse à l'intérieur de la population du Relevé Structurel. Cette intersection est calculée sur l'année n-1 et forme l'ensemble d'apprentissage pour imputer l'année n.

---

[5] Seules les relations « Epoux/se », « Partenaire enregistré-e », « Père/mère » et « Fils/fille » sont directement disponibles, les autres relations de la source « INFOSTAR » sont déduites de celles-ci.

[6] Pour les couples non-mariés ensemble ayant un enfant commun dans le même ménage.

[7] Dans la catégorie « apparenté » sont inclus notamment les oncles, tantes, cousins.

[8] Dans la catégorie « non-apparenté » sont inclus notamment les beaux-parents (parents de l'époux/se), les beaux-enfants (époux/se des enfants), les enfants du/de la concubin-e, le/la concubin-e du parent.

Résultats

Comparaison des distributions pour l'année 2018 : la distribution obtenue après imputation sur la population des ménages est comparée à la distribution estimée par le Relevé Structurel. La distance euclidienne entre ces deux distributions est de 1,92.

Tableau 2 : *Distributions des types de ménage pour l'année 2018 selon le Relevé Structurel et STATPOP*

Type de ménage	Estimation du Relevé Structurel	Typologie exhaustive (STATPOP)
Ménage d'une personne	35,69	35,69
Parent seul avec au moins un enfant de moins de 25 ans	4,64	5,24
Autre ménage de plusieurs personnes	7,59	8,01
Couple marié sans enfant	20,11	19,45
Couple en union libre sans enfant	6,53	7,51
Couple de même sexe sans enfant	0,61	0,25
Couple marié avec au moins un enfant de moins de 25 ans	22,12	20,87
Couple en union libre avec au moins un enfant de moins de 25 ans	2,70	2,96
Couple de même sexe avec au moins un enfant de moins de 25 ans	0,03	0,03

Estimation des erreurs individuelles : l'intersection entre l'échantillon des ménages du Relevé Structurel, et la sous-population des ménages imputés de la même année est utilisée pour comparer individuellement le type imputé. Ceci est possible, car les données du Relevé Structurel de l'année en cours ne sont pas utilisées dans la production du « Type de ménage » exhaustif ; elles servent ici d'ensemble de test pour le modèle d'imputation. Les poids du Relevé Structurel sont pris en compte dans le calcul de l'estimation de l'erreur individuelle sur la table des ménages.

Tableau 3 : *Estimation de l'erreur individuelle des imputations selon l'année*

Année	Pourcentage d'imputations incorrectes estimés sur le total des ménages
2018	1,45
2019	1,48
2020	1,47

Pour le moment, deux types d'erreurs représentent environ 76% du total des erreurs (soit env. 1.1% des 1.45%), il s'agit de la confusion des types : « Couple en union libre sans enfant » avec « Autre ménage de plusieurs personnes » (env. 56.5%), et « Couple de même sexe sans enfant » avec « Autre ménage de plusieurs personnes » (env. 19.5%). Cette confusion touche principalement les ménages de deux personnes : si les deux personnes sont de sexe opposé, l'algorithme a tendance à les considérer en union libre et si les deux personnes sont de même sexe, l'algorithme a tendance à les considérer comme colocataire. Cela implique une sous-estimation de la catégorie « Couple de même sexe sans enfant ». C'est notamment un des points qui va être retravaillé et qui améliorera de manière notable la qualité des imputations.

Conclusion

En conclusion, la variable type de ménage STATPOP peut être utilisée de manière relativement sûre ; en effet, 87% des types de ménage sont obtenus sur la base de données relevées et sur les 13% de ménages imputés, l'estimation de l'erreur de classification est faible (environ 1.45% de mauvaise attribution sur l'ensemble des ménages). Les premiers retours d'utilisateurs confirment l'utilité d'avoir cette variable de manière exhaustive. Quelques points sont encore à améliorer avant de pouvoir ajouter cette variable à la production courante de STATPOP, il s'agira notamment d'affiner l'imputation de certaines

catégories. Une documentation détaillée de la méthodologie est disponible sur le site de l'OFS (Typologie des ménages (STATPOP)).