

# RESUME

## RECENSEMENTS ET PUS

Le recensement de la population est un des ensembles de données les plus importants de la statistique officielle en Suisse. C'est un grand jeu de données dans tous les sens du terme: des données concernant tant les aspects démographiques que la situation professionnelle et sociale de la population sont récoltées sur l'ensemble de la Suisse. En outre, des informations sur les ménages et les conditions de logement peuvent également être analysées. Enfin, sa répétition tous les 10 ans permet de se faire une idée précise de l'évolution du pays. Il paraît donc extrêmement utile que les chercheurs et toute autre personne intéressée puissent avoir facilement accès à ces informations.

Les qualités des recensements constituent en même temps des difficultés pour l'analyse, à commencer par la taille du jeu de données et son exhaustivité. Même avec les moyens informatiques modernes, un jeu de données de plusieurs millions d'individus reste lourd à gérer. La difficulté essentielle n'est pourtant pas là, mais plutôt dans la préservation de la confidentialité des données: sachant qu'un recensement est exhaustif, il pourrait être possible de combiner les critères jusqu'à ce qu'une personne puisse être identifiée, ce qui est inacceptable aussi bien d'un point de vue éthique que légal. C'est ainsi que, pour pouvoir mettre à disposition ces données sans restrictions préalables, l'Office fédéral de la statistique (OFS) a décidé de tirer au hasard un échantillon représentatif de 5 % de la population à usage public (« Public Use Sample » ou PUS) en s'assurant que l'anonymat soit respecté.<sup>1</sup>

Les données décrites ici ne se limitent pas au seul recensement 2000, mais remontent jusqu'à celui de 1970. Cela rend possible une analyse de l'évolution – professionnelle, régionale, etc. – qui a caractérisé la Suisse durant cette période. La mise en parallèle de plusieurs recensements demande un travail spécifique pour harmoniser les données. En effet, un recensement, comme toute enquête, est un reflet de l'époque de sa réalisation: par exemple, les professions de l'informatique n'étaient pas les mêmes en 1970 qu'aujourd'hui, ce qui explique la nécessaire adaptation des nomenclatures ! Dans ce sens, et au-delà de la seule question technique, une réflexion est nécessaire sur les modalités de comparaison de données saisies à des moments différents.

Cette dernière remarque montre aussi très clairement l'importance de la documentation, qui doit refléter tous les aspects qui ont présidé à l'élaboration des données mais aussi tenir compte des buts pour lesquels les données ont été préparées et des modes de travail des usagers potentiels. Pour l'essentiel, cette description des données se base sur un gros travail de documentation technique élaboré par l'OFS ainsi que sur l'information originale sur les recensements que l'on peut trouver dans les nombreuses publications de l'OFS. La description que nous présentons ici se veut toutefois simple et accessible, tout en réunissant des informations issues de nombreuses publications.

---

<sup>1</sup> Outre l'anonymisation des données, où, conformément à la loi sur la protection des données, les indications nominales et les adresses sont éliminées, un deuxième traitement a été effectué pour éliminer les possibilités d'identification par recoupement. Ceci n'empêche bien sûr pas les données originales de pouvoir être utilisées, pour autant que les règles déontologiques et de confidentialité soient respectées et garanties par la signature d'un contrat spécifique.

## **A QUOI ET A QUI PEUVENT SERVIR DES PUS ?**

En bref, le but du travail présenté ici est de permettre à toute personne intéressée d'analyser les données du recensement dans le cadre d'un « modèle réduit », notamment dans les perspectives:

- de l'enseignement, par exemple, cours d'analyse statistique, autres cours ou travaux pratiques de recherche, afin d'introduire et de sensibiliser les étudiants à l'analyse du recensement;
- de recherches ou d'explorations en vue d'une recherche plus conséquente, par exemple analyses exploratoires ou préparatoires en vue d'encourager à débiter toute recherche par une étude préliminaire en partant de la statistique publique;
- de préparations de recherches qui nécessitent la connaissance de certaines caractéristiques de la population, de quotas, d'estimations de la taille de groupes spécifiques, préparation de classifications, essais d'agrégation etc.;
- de l'accès à des informations contextuelles afin d'enrichir les données recueillies dans le cadre d'autres recherches, ce qui implique de faire un lien sur les « niveaux géographiques » et les possibilités d'agrégation qu'ils représentent dans la documentation.

Quant au public cible, il s'agit de toute personne intéressée, quelle que soit son activité. On mentionnera notamment:

- les étudiants, encadrés dans des cours et séminaires;
- les doctorants et collaborateurs scientifiques des universités qui pourront se familiariser avec la statistique officielle, apprendre à formuler de manière plus précise les questions à aborder et à utiliser les données nécessaires pour y répondre;
- les collaborateurs des institutions privées et des administrations publiques qui se rendront ainsi mieux compte des possibilités du recensement et auront ainsi un accès facile à des données globales;
- les chercheurs étrangers, en particulier s'ils ont déjà la possibilité de travailler sur des PUS dans leur propre pays.

Un des intérêts majeurs des PUS ne réside paradoxalement pas dans les données que nous présentons ici. En effet, l'information a été réduite pour assurer l'anonymat, ce qui interdit ou limite certaines analyses. Cependant, en montrant la richesse d'information existante, en mettant en lumière les possibilités qu'offre le recensement, les PUS permettent de rendre plus aisé l'accès aux données du recensement: il est en effet plus facile d'adresser des demandes à l'OFS lorsque l'on connaît les variables disponibles et de poursuivre ensuite les analyses sur la base des données complètes.

## **QUELLE INFORMATION METTRE A DISPOSITION?**

Le recensement comprend en fait plusieurs jeux de données: sur les personnes, les ménages, les logements et les bâtiments. Pour simplifier l'exploitation, nous avons décidé de ne considérer ici que le fichier des personnes auquel nous avons joint des informations sur les ménages, les bâtiments et les logements. Dans le jargon technique, on parlera donc d'un « Public Use Sample » de personnes et non d'un « Public Use Sample » de ménages.

Pour chaque recensement, l'échantillon est obtenu par un tirage aléatoire simple des personnes dans le jeu de données complet, de manière à obtenir un échantillon de 5% de la population. Quant au contenu, nous nous sommes efforcés de retenir l'ensemble des informations disponibles et les seules réductions opérées l'ont été pour éviter qu'un répondant ne puisse être identifié. Dans ce

sens, nous avons systématiquement choisi de réduire le détail des réponses plutôt que de sélectionner un sous-ensemble de variables.

La philosophie qui sous-tend ce projet est donc celle d'un « produit d'appel »: en montrant l'ensemble des données à disposition, nous espérons que les utilisateurs seront tentés de recourir au recensement complet, ce qui impliquera cependant de s'engager contractuellement à préserver l'anonymat.

Concrètement, ce produit comprend trois types de contenus: des informations sur les recensements en Suisse et, surtout, les données ainsi recueillies; quelques indications sur les exigences de la protection des données et les moyens d'y faire face et, finalement, une documentation exhaustive des variables à disposition.