

Calcul de précision dans l'enquête emploi en continu.

Pascal ARDILLY¹, Guillaume OSIER²

¹ Insee, Unité Méthodes Statistiques

² Université Lyon I

L'enquête Emploi est une enquête individuelle menée chaque trimestre auprès d'environ 54 000 logements. Elle est basée sur un échantillonnage rotatif, conçu de manière à ce qu'un logement tiré un trimestre donné soit interrogé six trimestres de suite. L'échantillonnage est stratifié, à plusieurs degrés et s'appuie sur des sélections à probabilités inégales. Le logiciel POULPE développé à l'Insee, utilisable sous SAS-V8, a été utilisé pour estimer la précision de certaines statistiques issues de l'enquête Emploi, cela dans une approche transversale. Poulpe prétend être un outil "universel" en ce sens où il apparaît en mesure de traiter (presque) tous les plans de sondage, de prendre en compte la non-réponse et les procédures de redressement. Il est basé sur une approche analytique, c'est-à-dire qu'il s'appuie sur des formules de variance associées à chaque degré de tirage. Par ailleurs, on peut traiter n'importe quel estimateur complexe si on le linéarise au préalable.

L'enquête emploi a donné lieu à une stratification par région et par tranche d'unité urbaine. On a d'abord constitué des unités primaires à partir de découpages administratifs ou ayant un sens « démographique » (communes, groupes de districts, districts). Certaines unités ont alors été tirées avec une probabilité de sélection proportionnelle à leur taille, définie par le nombre total de logements (l'échantillon complet contient 2 554 unités primaires). Ces unités primaires ont ensuite été scindées en unités plus petites appelées "secteurs" (un secteur regroupe entre 120 et 240 logements), et un unique secteur a été sélectionné au hasard dans chaque unité primaire, proportionnellement à sa taille (en nombre total de logements). Chaque secteur a finalement été divisé en blocs d'une vingtaine de logements, appelés "aires". Un trimestre donné, une unique aire a été sélectionnée dans chaque secteur, par sondage aléatoire simple. Finalement, dans chaque aire tirée, tous les logements recensés en 1999 sont enquêtés tandis que certains logements neufs sont tirés par sondage aléatoire simple à partir des informations recueillies sur le terrain par les enquêteurs. Avec un tel plan de sondage, l'existence de tailles d'échantillon égales à un lors de certains tirages pose un problème pour estimer la précision. Pour y faire face, une technique de regroupement a priori d'unités primaires (appelée "collapse") a été mise en oeuvre et il a fallu procéder à des simplifications du plan de sondage.

La phase opérationnelle de POULPE s'appuie sur trois tables SAS : la première décrit précisément le plan de sondage, la seconde contient des informations indispensables au calcul des probabilités de sélection à chaque degré, et la troisième contient les données individuelles, à savoir les identifiants, les variables d'intérêt, les variables auxiliaires et une estimation des probabilités de réponse. En fait, les difficultés pratiques surviennent essentiellement durant la phase préparatoire - généralement assez laborieuse - des tables. Une fois cette phase réalisée, la mise en oeuvre proprement dite du logiciel est simple.

L'article présente les principales caractéristiques du logiciel Poulpe, le plan de sondage de l'enquête emploi et la façon dont il a été simplifié pour être exploité par le logiciel. Des calculs d'écart types et d'effets de sondage sont également fournis, dans une approche transversale.