



HAL
open science

Pondérations du panel ELIPSS

Blazej Palat

► **To cite this version:**

| Blazej Palat. Pondérations du panel ELIPSS. 2023. hal-04329785

HAL Id: hal-04329785

<https://sciencespo.hal.science/hal-04329785v1>

Submitted on 7 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Pondérations du panel ELIPSS

Guide utilisateur

Introduction

La force du dispositif ELIPSS réside dans la facilité avec laquelle ce panel probabiliste peut être mobilisé pour la recherche par questionnaire ainsi que dans sa dimension longitudinale. Ces avantages incontestables ont un prix, notamment la taille relativement modeste de l'échantillon et, par conséquent, une représentativité imparfaite sur le plan de quelques caractéristiques socio-démographiques importantes. Lorsqu'on souhaite pallier ces désavantages par la pondération, les individus dont les caractéristiques sont rares en termes de proportion par rapport à la population de référence se voient attribuer des poids excessivement élevés, devenant ainsi trop influents dans les estimations.

Deux pistes d'action ont été mises en œuvre afin de limiter l'impact de ce phénomène. Des calages pénalisés d'une part, résultant d'un redressement partiel par rapport aux estimations de référence (marges de l'Enquête Emploi en continu). D'autre part, des troncatures unilatérales des poids les plus élevés. L'utilisateur des pondérations mises à disposition doit donc tenir compte du redressement imparfait, notamment sur le plan du niveau de diplôme des répondants, les catégories BAC – BAC+2 étant surreprésentées, alors que celles inférieures au BAC sous-représentées dans le panel.

L'objectif du présent guide est conforme à l'approche adoptée pour la mise à disposition des pondérations depuis le début du dispositif. Le fichier de données contient les poids résultant des deux phases de correction de la non-réponse totale : au recrutement et à l'enquête en question. La version précédente de la pondération (avant le rafraîchissement du 2023) est également incluse dans le fichier. De plus, nous proposons à l'utilisateur un outil annexe, destiné à lui faciliter le calcul de pondérations de la seconde phase de la correction de la non-réponse totale en toute autonomie, en modifiant à sa guise les paramètres essentiels de ce calcul.

Pondérations mises à disposition dans le jeu de données

Poids des répondants à l'enquête – *seconde phase de correction de la non-réponse totale*

PDSPLT_[NOM DE L'ENQUETE]

Nom de la variable correspondant à la pondération propre à l'enquête, s'appliquant au panel rafraîchi en 2020. A utiliser en cas d'appariements de variables provenant d'enquêtes réalisées dans la période 2021-2022 (avant l'intégration de la vague de rafraîchissement de 2023).

POIDS_[NOM DE L'ENQUETE]

Nom de la variable correspondant à la pondération propre à l'enquête, s'appliquant au panel rafraîchi en 2023. A utiliser pour l'enquête en question y compris en cas d'appariements de variables d'enquêtes réalisées à partir de 2023.

Les poids correspondant à la pondération enquête servent à pondérer les résultats. *PDSPLT_* contiennent des valeurs manquantes (panélistes recrutés lors de la vague de rafraîchissement de 2023). Afin de réduire leur dispersion, les poids calés sont tronqués au seuil de 5 % des poids les plus élevés.

Poids initiaux – *première phase de correction de la non-réponse totale*

PDSPLT_INIT

Nom de la variable correspondant à la pondération initiale du panel rafraîchi en 2020.

POIDS_INIT

Nom de la variable correspondant à la pondération initiale du panel rafraîchi en 2023.

Les poids initiaux ne servent pas à pondérer les résultats. Ils sont fournis pour les utilisateurs qui souhaitent réaliser eux-mêmes les calages (et donc prendre en charge eux-mêmes la seconde phase de correction de la non-réponse). Le code R (R CORE TEAM, 2021), faisant appel aux fonctionnalités du package `icarus` (REBECQ, 2016), utile dans ce but est fourni sous forme d'un script. Comme dans le cas des poids propres à l'enquête, les poids `PDSPLT_` contiennent des valeurs manquantes pour les panélistes recrutés lors du rafraîchissement du 2023.

Script annexe

Ce script contient un ensemble de fonctions « wrapper » permettant d'effectuer facilement des calages et des troncatures de poids avec des valeurs par défaut correspondant aux routines appliquées dans la production des fichiers d'enquête ELIPSS. Dans la suite de ce document, nous présentons la fonction `calageELIPSS`, les autres fonctions ne nécessitant pas de documentation détaillée. Elle dépend des packages `icarus`, `survey` et `ggplot2`.

Usage

```
calageELIPSS(  
data,  
gap=0.5,  
nomVar="Utilisateur",  
pdsplt=FALSE,  
sumTotPop=FALSE,  
troncatures=seq(0.99,0.8,-0.01),  
vars=c("cal_tuu","cal_age","cal_dipl","cal_sexe","cal_zeat"),  
catTestVars=NULL,  
contTestVars=NULL)
```

Arguments

data	<code>data.frame</code> . En l'occurrence, un jeu de données ELIPSS (avec les poids initiaux et les variables de calage).
gap	numeric, length 1. Écart fixé entre les coefficients de calage pour le calage pénalisé.
nomVar	character, length 1. Nom de la variable correspondant à la pondération créée par l'utilisateur.
pdsplt	logical, length 1. Choix d'utiliser la pondération <i>pdsplt</i> .
sumTotPop	logical, length 1. Choix d'utiliser la pondération non-normalisée (sommant au total de la population).
troncatures	numeric. Choix de niveaux de troncature des poids créés à ajouter dans <code>data</code> .
vars	character. Choix de variables de calage.
catTestVars	character. Variables (catégorielles) à utiliser pour illustrer les écarts cumulés entre les tri pondérés par des poids tronqués et non-tronqués.
contTestVars	character. Variables (continues) à utiliser pour le calcul de l'erreur quadratique moyenne suite à différents niveaux de troncature.
...	Autres arguments de la fonction <code>icarus::calibration</code> à l'exception de <code>marginMatrix</code> , <code>colWeights</code> , <code>popTotal</code> et <code>pct</code> .

Valeur

Si l'argument `troncatures` n'est pas `NULL`, objet de classe `list` de 5 éléments :

- `Descriptives` Liste de matrices contenant des statistiques descriptives des pondérations tronquées.
- `Ecart` Matrice des écarts entre les tris des variables `catTestVars` pondérés par des poids tronqués à différents seuils et les tris pondérés par des poids non-tronqués ou `NULL`.
- `EQMs` Matrice d'erreurs quadratiques moyennes pour les variables `contTestVars` à chaque niveau de troncature ou `NULL`.
- `Graph.ecarts` Graphique représentant les écarts entre les tris des variables `catTestVars` pondérés par des poids tronqués à différents seuils et les tris pondérés par des poids non-tronqués ou `NULL`.
- `data` Jeu de données enrichi de la nouvelle pondération avec les troncatures le cas échéant.

Sinon, juste un objet de classe `data.frame` enrichi de la nouvelle pondération.

Détails

Cette fonction est un simple « wrapper » de `icarus::calibration` destiné à caler les pondérations initiales des jeux de donnée ELIPSS sur la base des marges de l'Enquête Emploi en continu de 2019 et 2021 de l'Insee. Il offre une fonctionnalité additionnelle (fonction troncature) permettant d'effectuer une troncature unilatérale (limitant le nombre d'unités influentes) en appliquant une méthode détaillée dans le document de travail de MARQUIER (2010). La valeur de `gap` est fixé à 0.5 par défaut, ce qui correspond à la valeur de cet argument appliquée systématiquement dans la seconde phase de correction de la non-réponse totale (calage pénalisé). Alors que `costs` est NULL par défaut, en réalité, lorsque `gap` n'est pas spécifié comme NULL `costs` passe automatiquement à `c(1,1,1,1,1)` lorsque `pdsplt` est FALSE et à `c(200,100,1,1,1)` lorsque `pdsplt` est TRUE. Le jeu de donnée produit par la fonction est écrémé des panélistes recrutés en 2023 (non-concernés par la pondération `pdsplt`) lorsque `pdsplt` est TRUE.

Exemples

```
### 'data' est un jeu de données ELIPSS
#
# Une fois le fichier 'fonctions.R' placé avec les autres éléments (en particulier
# 'data') dans le répertoire de travail

source('fonctions.R')

# calage de la pondération initiale poids_init avec un écart de coefficient
# augmenté et nouvelles valeurs de coûts, sans aucune troncature des poids

calageELIPSS(data, gap=1, costs=c(1,1,100,50,500), troncatures=NULL)

# calage de la pondération initiale pdsplt_init, sommant au total de la population
# de référence avec la méthode 'raking' (non-pénalisé, calage précis) et avec les
# variables de calage illustrant l'impact des troncatures

calageELIPSS(data, gap=NULL, pdsplt=TRUE, sumTotPop=TRUE, method='raking'
, catTestVars=c("cal_age", "cal_dipl", "cal_sexe", "cal_tuu", "cal_zeat"))

# calage de la pondération initiale poids_init avec la méthode 'raking'
# (non-pénalisé, calage précis) sans la variable de calage 'cal_dipl' (niveau de
# diplôme)

calageELIPSS(data, vars=c("cal_age", "cal_sexe", "cal_tuu", "cal_zeat")
, gap=NULL, method="raking")
```

Références

- MARQUIER, R. (2010). *Echantillonnage, apurements et redressement de la non réponse dans l'enquête IAD*. Série sources et méthodes, Direction de la recherche, des études, de l'évaluation et des statistiques (DREES).
- R CORE TEAM. (2021). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- REBECQ, A. (2016, octobre). *Icarus, un package R pour le calage sur marges et ses variantes*. 9^e Colloque International Francophone sur les Sondages, Gatineau, Canada. http://paperssondages16.sfds.asso.fr/submission_54.pdf