Vous avez été contacté par Médiamétrie ? En savoir plus

Redressement d'échantillon : principe et méthodes

Le redressement d'échantillon consiste à corriger les déformations de l'échantillon pour rapprocher sa structure de celle de la population étudiée. À travers le calage sur marges ou des algorithmes comme CALMAR et Ridge, il contribue à limiter les biais, renforcer la précision statistique et fiabiliser les indicateurs issus des enquêtes.

Qu'est-ce qu'un échantillon représentatif ?



Un sondage n'est pas un recensement. Le sondage est une méthode qui consiste à interroger une partie de la population – un échantillon - et à l'analyser afin d'extrapoler les résultats de l'échantillon à l'ensemble de la population. Toutes les enquêtes par sondage sont sujettes à différents types d'erreur, dont la principale est l'erreur d'échantillonnage. Celle-ci est due à l'aléa du tirage de l'échantillon.

Pour maîtriser l'erreur d'échantillonnage et améliorer la précision des résultats de l'enquête, on utilise des informations auxiliaires. Il s'agit d'informations recueillies dans l'enquête et dont on connaît les valeurs dans la totalité de la population. Ces informations peuvent intervenir à plusieurs niveaux : lors de la conception du plan de sondage avec la stratification ou les quotas, ou lors du traitement avec le redressement.



Pourquoi le redressement d'échantillon est-il indispensable ?

Un échantillon ne peut pas être une copie conforme de la population étudiée. Si l'on publie les résultats « bruts », on obtient une image déformée de la réalité.

Le redressement (ou *calage*) consiste à corriger a posteriori les déformations de l'échantillon sur un certain nombre de critères, en attribuant à chaque répondant un poids qui reflète ce qu'il représente dans la population. Concrètement, un individu appartenant à une catégorie sous-représentée recevra un poids supérieur à 1 (sa réponse « comptera davantage »), et inversement pour une catégorie surreprésentée.

L'enjeu est double :

- Fiabilité : garantir que les indicateurs publiés ne sont pas biaisés du fait de la structure de l'échantillon.
- Précision : un redressement sur des critères corrélés aux variables d'intérêt réduit la variance des estimations et améliore la précision des résultats.

En poursuivant votre navigation, nous vous informons qu'un cookie technique de session sera déposé afin de faciliter votre navigation sur notre site.
Plus d'infos

Oui, je suis d'accord

Non, merci



exemple les données de l'Insee pour les critères socio-démographiques. L'objectif est de faire en sorte que la structure de l'échantillon un fois pondéré coïncide avec ces données de cadrage sur chaque critère retenu.

Cas simple : un seul critère

Si l'on souhaite redresser l'échantillon sur un seul critère, le genre par exemple, alors il s'agit d'une simple règle de trois. Par exemple, si les femmes représentent 52 % de la population mais seulement 40 % de l'échantillon, chaque femme reçoit un poids de 52/40 = 1,30 et chaque homme un poids de 48/60 = 0,80.

Cas classique : plusieurs critères - le calage sur marges

En pratique, on redresse simultanément sur plusieurs critères (sexe, âge, CSP, région, type d'habitat...). On ne connaît en général que les distributions marginales sans disposer de la totalité des distributions croisées. La méthode du calage sur marges consiste alors à trouver des poids qui respectent toutes ces marges à la fois, tout en restant aussi proches que possible des poids initiaux.

Il s'agit d'un problème d'optimisation sous contraintes : on minimise une fonction de distance entre les poids de sondage et les poids de calage, sous la contrainte que les marges soient exactement respectées.

Quelles sont les principales méthodes de redressement d'échantillon ?

La méthode itérative RAS : ajuster progressivement les poids de l'échantillon

La méthode la plus intuitive est le RAS (Raking Adjustment System ou *iterative proportional fitting*). Le principe est simple :

1. On ajuste d'abord les poids pour que la marge du premier critère (par exemple le sexe) soit exacte.

2. Puis on réajuste pour que la marge du deuxième critère (l'âge) soit respectée — ce qui peut légèrement déformer la première marge.

3. On recommence sur le premier critère, et ainsi de suite, en boucle.

À chaque passage, les poids sont multipliés par le rapport entre la marge théorique et la marge estimée dans l'échantillon pondéré :

où M_j est le total de population dans la catégorie j et le dénominateur est le total pondéré courant dans cette catégorie.

L'algorithme converge généralement en quelques itérations. Sa fonction de distance sous-jacente est G(x)=xlog(x)-x+1, ce qui garantit des poids strictement positifs. En revanche, il peut produire des poids très élevés si les écarts entre l'échantillon et la population sont importants.

La macro CALMAR de l'INSEE, une méthode de référence pour le calage sur marges

Développée par l'Insee, la macro CALMAR (CALage sur MARges) formalise le calage comme un problème d'optimisation : trouver les poids w_k minimisant une fonction de distance G par rapport aux poids de sondage d_k, sous les contraintes de calage :

Le système est résolu par l'algorithme de Newton. La macro CALMAR propose cinq fonctions de distance asymptotiquement équivalentes : le choix entre elles est en pratique guidé par l'allure de la distribution des poids et par la volonté d'éviter des valeurs extrêmes.

L'approche Ridge : une méthode plus souple quand les critères sont nombreux

Lorsque les critères de calage sont nombreux ou partiellement colinéaires, le système d'équations peut devenir instable ou ne pas converger. L'approche Ridge (ou régularisation L2) offre une solution en ajoutant un terme de pénalisation à la matrice du système.
Dans le cadre du calage, au lieu de résoudre $\beta^ = (X' X)^{-1} X' Y$, on résout :
$\beta^_{ridge} = (X' X + \lambda I)^{-1} X' Y$
Le paramètre $\lambda>0$ régularise la matrice X' X en la rendant mieux conditionnée : les valeurs proches de zéro sont « relevées », ce qui stabilise l'inversion. Le prix à payer est un léger biais sur les coefficients, compensé par une réduction substantielle de la variance.
En pratique, cette approche est utile lorsque :

- Le nombre de critères de calage est élevé par rapport à la taille de l'échantillon ;
- Les informations auxiliaires sont elles-mêmes issues d'enquêtes et sujettes à des marges d'erreur ;

On souhaite un calage « souple » qui ne respecte pas exactement les marges mais produit des poids plus stables.

À retenir sur le redressement d'échantillon

Le redressement n'est pas une opération cosmétique : bien conduit, il améliore significativement la précision des estimations. Le choix de la méthode dépend du contexte, du nombre de critères, de la qualité des données de cadrage et des contraintes opérationnelles.

Notre newsletter pour suivre nos dernières
Inscrivez-

études, analyses et actualités

vous

Retrouvez toutes nos actualités sur les réseaux sociaux

Mediametrie
High trust data

+33 1 47 58 97 58
70 rue Rivay 92532
Levallois-Perret Cedex

Amiens - Sophia Antipolis - Mont Saint Aignan
Voir tous les sites

Contactez-nous

A propos

- Société à Mission
- Organisation
- Partenariats
- Nos engagements

Audiences

- TV
- Radio
- Internet
- Vidéo
- Podcasts
- Outre-Mer
- International

Solutions

- MediaEar
- MediaMat
- MediaWeb
- MediaWatch
- MediaTrend
- MediaFrom
- MediaReach

Expertises

- Expertise Méthodologique
- Publications
- Outils statistiques

Comprendre

- La mesure d'audience
- M Insights : nos analyses
- Panélistes et Interviewés
- Mots de la Mesure : glossaire
- FAQ

Presse **Actualités** **Carrières**

©2026 Accessibilité CGU CGV Mentions légales Confidentialité Gestion cookies Données personnelles Plan du site