

# MESURE D'AUDIENCE GLOBALE DES MARQUES DE PRESSE : ENJEUX METHODOLOGIQUES A L'HEURE DU NUMERIQUE

Flavien Alleaume<sup>1</sup> & Gaël Crochet<sup>2</sup>

<sup>1</sup> *Médiamétrie, 70 rue Rivay, 92532 Levallois-Perret Cedex, falleaume@mediametrie.fr*

<sup>2</sup> *Médiamétrie, 70 rue Rivay, 92532 Levallois-Perret Cedex, gcrochet@mediametrie.fr*

**Résumé.** Les marques de presse ne sont aujourd'hui plus simplement lues sur leurs supports papier, mais une partie importante et croissante de leur lectorat se tournent vers les versions numériques de ces titres. L'étude ONE Global, copropriété d'Audipresse et de Médiamétrie, permet de répondre au besoin du marché de disposer d'une étude unique de référence pour mesurer la couverture des marques de presse sur tous les supports. Une technique de fusion d'enquêtes a été mise en place pour rapprocher l'étude ONE, et les panels Internet Mobile et Médiamétrie NetRatings, références respectives des mesures d'audience de la Presse, d'Internet mobile et d'Internet fixe (i.e. sur ordinateur). Cette méthode permet de préserver les duplications intra-médias de chacune des études de référence, tout en ne dégradant pas les duplications inter-médias, portées par l'étude ONE. De nouvelles méthodologies ont été mises en œuvre à chaque stade du projet (stratification, distance, jumelage) sous le contrôle scientifique du CESP (centre d'étude des supports publicitaires).

**Mots-clés.** Données d'enquête, Fusion, Stratification, Jumelage.

## 1 Introduction

En partenariat avec Médiamétrie, Audipresse a souhaité lancer en Avril 2014 une mesure globale de l'audience des titres de presse. Les évolutions des modes de consommation du média rendent nécessaire une méthodologie prenant en compte toutes ses composantes. En effet, le marché de la presse subit actuellement une mutation de ses modes de lecture, redéfinissant par la même les enjeux de la mesure d'audience.

Ainsi, la volonté du projet ONE Global est de réunir en une seule étude trois types de consultation de la presse : le papier, l'Internet Fixe, et l'Internet Mobile, chacune respectivement issues des trois études référentes sur le marché : l'étude ONE, le panel Médiamétrie NetRatings et la mesure d'audience d'Internet Mobile.

L'un des points de complexité consiste à rapprocher des études aux caractéristiques différentes. Ainsi, ONE est une enquête ponctuelle déclarative renouvelé par moitié 2 fois par an. A l'inverse, l'approche dans les 2 enquêtes Internet repose sur 2 panels de mesure d'audience automatique.

Les travaux présentés dans cet article reprennent les éléments méthodologiques mis en œuvre au cours de la construction du projet, mené en collaboration avec le CESP, la direction des études d'Audipresse et la Direction Analyses et Méthodes Scientifiques de Médiamétrie.

## 2 Fusion d'enquêtes

Les deux enquêtes Internet (fixe et mobile) de Médiamétrie nous permettent, pour chaque panéliste,

de connaître s'il s'est ou non connecté à un site de presse donné au cours du mois de mesure. L'enquête ONE, quant à elle, contient pour chaque titre de presse et pour chaque individu les dates de dernière connexion à toutes les versions fixes et mobiles des titres de presse.

On dispose donc de variables de pont entre l'enquête ONE et les panels Internet fixe et mobile permettant de réaliser une fusion d'enquêtes. L'objectif est ici de préserver d'une part les duplications inter-médias, portées par les réponses apportées par les individus ONE (seuls individus à avoir décrit leur comportement sur chacune des versions, papiers ou numériques, des différentes marques de presse), et d'autre part les duplications intra-médias, portées de manière référente par les enquêtes respectives.

Usuellement, le processus de fusion d'enquêtes se déroule en trois étapes :

1. Stratification : création de groupes d'individus au sein desquels seront faits les jumelages. Les jumeaux seront donc « parfaits » sur les critères de stratification.
2. Définition d'une distance entre les individus : détermination des critères permettant d'évaluer la proximité entre individus.
3. Jumelage : c'est lors de cette étape qu'est défini le mode de sélection du jumeau.

## 2.1 Stratification

L'objectif de cette étape est double. Nous cherchons tout d'abord à respecter certains critères socio-démographiques majeurs (sexe, âge ...) entre le receveur et le donneur, notamment dans le cadre d'une utilisation commerciale ultérieure. D'autre part, nous souhaitons constituer des strates homogènes de comportement afin de s'assurer d'attribuer aux receveurs des jumeaux qui leur ressemblent en termes de comportement.

Pour identifier ces strates, on cherche à caractériser le surf des panélistes. Sur le fixe comme sur le mobile on utilise des variables résumant cette information. On distingue ainsi 15 variables relatant le volume de surf sur 15 catégories de sites.

Chacune de ces variables est discrétisée en terciles (hors jamais), donnant ainsi naissance à une variable qualitative à 4 modalités: un groupe pour les non-concernés, puis une division en 3 groupes de tailles égales selon que le panéliste est un petit, moyen ou gros consommateur.

On dispose dès lors d'une soixantaine de variables dichotomiques caractérisant le surf du panéliste, et utilisables via une Analyse des Correspondances Multiples (ACM).

L'ACM mise en place sur les données Fixe comme sur les données Mobile nous donne un premier axe factoriel qui réunit, via la mesure corrigée de Benzécri de l'inertie, plus de 80% à lui seul de l'information.

L'idée sur laquelle repose la méthodologie est la suivante :

Puisque le premier axe représente l'essentiel de l'information contenue dans le nuage des individus, si une variable socio-démographique est significativement discriminante au regard de cet axe, alors elle l'est également quant à la consommation de l'Internet (fixe ou mobile selon le cas).

Ainsi, nous avons placé en variables illustratives les variables socio-démographiques dont nous disposons lors de l'ACM. Ces variables ne participent donc pas à la formation des axes, mais sont projetées dessus.

Nous pouvons alors tester la significativité de la coordonnée de la modalité à laquelle on s'intéresse. Cette étape possède un double intérêt. Elle permet d'une part d'identifier les variables socio-démographiques qui sont statistiquement significatives sur le premier axe, et d'autre part de projeter la variable de stratification sur cet axe et de vérifier que toutes les modalités de cette variable sont significatives. Autrement dit, si une modalité de la variable de strate n'était pas significative, cela voudrait dire que les individus de la strate se comportent de la même façon que l'ensemble de la population sur l'axe factoriel retenu.

On cherche donc une stratification respectant les critères suivants :

- Chacune des modalités est statistiquement discriminante sur le premier axe,
- Elle est construite à partir de variables socio-démographiques elles-mêmes significatives,
- Le nombre d'individus dans chacune des strates y est suffisamment important (au moins 200 sur l'enquête donneuse).

La méthodologie permettant de tester l'effet discriminant d'une variable de stratification n'assure pas que les strates aient toutes des comportements différents, et que certaines pourraient être regroupées au sein d'une même strate. Cependant, rajouter un critère socio-démographique supplémentaire pour obtenir deux strates au lieu d'une (sous réserve d'avoir toujours au moins 200 individus dans chacune de ces deux sous-strates) permet de contrôler une variable socio-démographique supplémentaire dans le jumelage.

En pratique, on cherche tout d'abord à introduire une variable de strates d'initialisation (recoupant le sexe, l'âge et la CSP) comme variable illustrative. Puis, à l'aide des variables socio-démographiques statistiquement discriminantes, on procède itérativement à divers recoupements, jusqu'à obtenir une stratification permettant de répondre aux trois critères précédemment définis.

## **2.2 Distance entre individus**

Concernant l'enquête receveuse, on dispose de dates de dernière connexion à Internet, échelonnées comme suit :

- ⇒ Hier
- ⇒ Il y a moins de 8 jours
- ⇒ Il y a 8 à 15 jours
- ⇒ Il y a 15 à 30 jours
- ⇒ Il y a plus de 30 jours
- ⇒ Jamais

Par nature, les données de mesure d'audience Internet ne peuvent permettre de déterminer quelle est la « date de dernière connexion » du panéliste. La seule information dont nous disposons est, si oui ou non le panéliste a visité l'entité considéré au cours du dernier mois. On se propose donc d'utiliser cette information.

Ainsi, si l'on regroupe la donnée ONE pour la rendre comparable à la donnée de mesure d'audience Internet on obtient une variable en 3 modalités :

- ⇒ Moins de 30 jours
- ⇒ Plus de 30 jours
- ⇒ Jamais

Le principe consiste à ne pas dichotomiser la variable de date de dernière connexion de ONE (« Connecté il y a moins d'un mois » / « Non connecté il y a moins d'un mois »), pour garder l'information des panélistes qui ont déclaré « Il y a plus de 30 jours » comme date de dernière connexion. La logique est ici de considérer que, par défaut, il est préférable de jumeler un « Il y a plus de 30 jours » avec un visiteur dernier mois sur les mesures Internet, qu'un « Jamais » avec un visiteur dernier mois.

Dès lors, on décide de définir une « distance » dite « scorée » pour quantifier l'éloignement entre les individus receveurs et donneurs.

Le principe de la « distance » scorée est le suivant.

Soit une matrice de scores de type :

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}$$

Avec :

- $x_{11}$  le score attribué à « Receveur : « Moins de 30 jours ONE / Donneur : Moins d'un mois Internet »
- $x_{12}$  le score attribué à « Receveur : Moins de 30 jours ONE / Donneur : Non-audience Mois Internet »
- $x_{21}$  le score attribué à « Receveur : Plus de 30 jours ONE / Donneur : Moins d'un mois Internet »
- $x_{22}$  le score attribué à « Receveur : Plus de 30 jours ONE / Donneur : Non-audience mois Internet »
- $x_{31}$  le score attribué à « Receveur : Jamais ONE / Donneur : Moins d'un mois Internet »
- $x_{32}$  le score attribué à « Receveur : Jamais ONE / Donneur : Non-Audience mois Internet »

En procédant avec des jeux de paramètres  $x_{rd}$  adéquats ( $r=1,2$  ;  $d=1,2,3$ ), on s'offre la possibilité de considérer différemment tel ou tel type de jumelage et donc privilégier des jumeaux représentant une audience sur le titre à des jumeaux jamais lecteurs du titre. Il est en effet préférable de marier des individus parce qu'ils sont fréquentants d'un site ou utilisateur d'une application (qui sont plutôt rares) plutôt que parce qu'ils ne sont pas fréquentants d'un site ou d'une application (qui sont très nombreux).

On a ainsi pour chaque entité  $T$  considérée dans la distance (variables de pont), et pour chaque couple de jumeaux  $(i,j)$  un score basé sur les réponses de  $i$  et  $j$  relativement à la variable de pont.

$$Score_T(i, j) = \text{Nb points résultant de la règle définie ci-dessus}$$

Pour chaque couple d'individus  $(i, j)$ , on agrège les scores de chaque entité  $T$  pour obtenir un score global.

$$Score(i, j) = \sum_T \frac{Score_T(i, j)}{\gamma_T}$$

Avec  $\gamma_T$  l'audience du titre sur l'enquête référence. Le fait de pondérer le score attribué par l'inverse de l'audience du titre a pour objectif de permettre de redonner de l'importance aux petits titres dans le processus de jumelage. Sans cette pondération, la rareté de l'évènement de lecture de ces titres aurait conduit trop souvent à ce qu'un visiteur du site soit attribué à un non-visiteur ONE.

Comme on préfère les distances à minimiser plus que les scores à maximiser, on a défini la « distance » (comprise entre 0 et 1) entre  $i$  et  $j$  comme suit :

$$Distance(i, j) = 1 - \frac{1}{\sum_T \frac{1}{\gamma_T}} \sum_{i,j} \frac{Score(i, j)}{Max(Score)}$$

Avec  $Max(Score)$  équivalent au maximum des scores possibles (cas où les 2 individus lisent tous les titres), score de référence identique strate à strate.

A noter : le terme distance est ici un abus de langage : cette métrique ne respecte pas les propriétés mathématiques d'une distance.

### 2.3 Jumelage

Une fois la matrice de « distance » définie strate par strate, il convient d'attribuer à chaque receveur un unique donneur Internet fixe (et/ou mobile selon ses équipements). La méthode retenue est celle dite des « voisinages réciproques ».

Celle-ci revient à considérer comme jumeaux les couples (*receveurs*, *donneurs*) pour lesquels l'attraction (la distance faible) est réciproque. Concrètement, cela revient à rechercher au sein de la matrice de distances entre les individus *receveurs* (les lignes de la matrice) et les individus *donneurs* (les colonnes de la matrice) le couple de distance minimale, définissant ainsi le premier couple de jumeaux. On réitère cette opération à partir de la matrice de distance de l'itération précédente en ayant supprimé la ligne correspondant au *receveur* retenu lors de cette itération (un receveur ne peut servir qu'une seule fois) et en pénalisant les distances associées au donneur correspondant. La fonction de pénalisation est la suivante,  $j$  désignant le donneur à l'itération courante qui a été déjà utilisé  $n$  fois au cours des itérations précédentes :

$$\forall i \in \text{Receveurs} \quad d_{n+1}(i, j) = 1 - \left(1 - d_n(i, j)\right)^{\frac{n+\lambda}{n}}$$

L'idée de cette fonction est de faire tendre vers 1 les distances des individus au fur et à mesure qu'ils sont sélectionnés comme donneurs. A noter que  $d_0$  est la matrice de distance initiale normée à 1.

Le paramètre  $\lambda$  est à ajuster en fonction de l'importance que l'on souhaite donner à la pénalisation, c'est-à-dire de la volonté que l'on a de limiter les répliques. La méthode que nous avons utilisée

permet également de définir un seuil de répliques maximum, au-delà duquel le donneur est simplement retiré de la base des potentiels donneurs.

## **Bibliographie**

- [1] Lebart, L., Morineau, A. and Piron, M. (1995), *Statistique Exploratoire Multidimensionnelle*, ed. Dunod, 123-127.
- [2] Santini, G. (2003), *Mathematical models & Methods for Media Research*, G.S. IT Services, Paris.
- [3] Vanheuverzwyn A. (2009), *Fusion des études de référence de la mesure d'audience des médias Radio, TV, Internet et Presse*, X<sup>èmes</sup> Journées de Méthodologie Statistique de l'Insee, Paris.