

Contenus médiatiques et pratiques d'information des Français

Réalisé par

Carfantan Adrien & Hardy Pierre

Université de Tours
Master Économie de l'Entreprises et des Marchés



Année universitaire 2025–2026

Contents

1	Introduction	2
1.1	Enjeux et objectifs du projet	2
1.2	Présentation des bases de données	2
2	Partie I – ACP	4
2.1	4
2.2	4
3	Partie II – AFC	5
3.1	Objectif et validation statistique	5
4	Partie III – ACM	6
4.1	6
4.2	6
5	Partie IV – Clustering	7
5.1	7
5.2	7
6	Conclusion générale	8
7	Annexe	9

1 Introduction

1.1 Enjeux et objectifs du projet

Ce projet a été réalisé dans le cadre du cours de Data Mining, l'enjeu principal de celui ci réside dans la nature même des données étudiées, caractérisées par un nombre élevé de variables et par une forte hétérogénéité de l'information. L'analyse directe de ces données s'avère peu lisible, ce qui rend nécessaire le recours à des méthodes statistiques capables de synthétiser l'information tout en conservant les structures essentielles. L'objectif est ainsi de faire apparaître des régularités, des oppositions et des proximités entre les variables et les individus.

Le projet poursuit plusieurs objectifs. Il s'agit d'abord de mettre en évidence les grandes dimensions qui structurent la couverture thématique de l'information télévisée, afin de mieux comprendre l'organisation des contenus diffusés dans les journaux télévisés. Il vise également à analyser les pratiques d'information des individus, en étudiant les relations entre leurs caractéristiques sociales, leur intérêt pour l'actualité, leurs sources d'information et leur perception de la fiabilité des médias.

Pour répondre à ces objectifs, différentes méthodes d'analyse multivariée ont été mobilisées. Une Analyse en Composantes Principales a d'abord été appliquée afin d'identifier les principaux axes organisant les thématiques de l'information. Une Analyse Factorielle des Correspondances a ensuite permis d'étudier les liens entre les catégories d'âge et l'usage des réseaux sociaux. Enfin, une Analyse des Correspondances Multiples a été réalisée afin d'analyser conjointement plusieurs variables qualitatives et de dégager des profils d'individus en fonction de leurs pratiques et perceptions de l'information.

Cette approche est complétée par une classification des individus fondée sur les résultats de l'ACM, dont l'objectif est de regrouper les répondants en classes homogènes et interprétables. L'ensemble de ces analyses permet ainsi de proposer une lecture synthétique des données et d'apporter des éléments de compréhension sur la diversité des comportements face à l'information et aux médias dans le contexte français contemporain.

1.2 Présentation des bases de données

Le travail réalisé dans ce projet repose sur l'exploitation de deux bases de données distinctes, issues d'organismes publics français, et portant toutes deux sur la thématique de l'information et des médias. Bien que différentes par leur nature et leur structure, ces bases sont complémentaires et permettent d'aborder le sujet sous deux angles : celui de la production de l'information et celui de sa réception par le public.

La première base de données est fournie par l'Institut National de l'Audiovisuel (INA). Elle recense, sur une longue période, les sujets traités dans les journaux télévisés français, en distinguant plusieurs thématiques telles que la politique, l'économie, la santé, la culture, le sport ou encore les faits divers. Chaque observation correspond à un couple date-chaîne et indique le nombre de sujets diffusés pour chaque thème. Cette base est essentiellement quantitative et se prête naturellement à une Analyse en Composantes Principales, dont l'objectif est de dégager les principales dimensions structurant la couverture médiatique des journaux télévisés.

La seconde base de données provient d'une enquête menée par l'ARCOM intitulée « Les Français et l'information ». Elle repose sur des données d'enquête collectées auprès d'un large échantillon de la population française et contient principalement des variables qualitatives. Ces variables décrivent à la fois des caractéristiques socio-démographiques (sexe, âge, niveau d'études, situation professionnelle) et des pratiques liées à l'information, telles que l'intérêt pour l'actualité, les sources utilisées pour se forger une opinion ou encore la perception de la fiabilité de l'information. La richesse et la diversité de ces variables rendent cette base particulièrement adaptée à des méthodes d'analyse factorielle pour données qualitatives, notamment l'Analyse des Correspondances Multiples.

L'utilisation conjointe de ces deux bases de données permet d'adopter une approche globale du phénomène étudié. D'un côté, la base de l'INA offre une vision structurée des contenus informationnels diffusés par les médias audiovisuels ; de l'autre, la base de l'ARCOM permet d'analyser la manière dont ces contenus sont

perçus, consommés et évalués par les individus. Ce double regard constitue un atout majeur du projet et justifie le recours à des méthodes statistiques variées afin d'exploiter au mieux l'information contenue dans les données.

2 Partie I – ACP

2.1

2.2

3 Partie II – AFC

3.1 Objectif et validation statistique

Afin de completer notre analyse, passons a l'étude des liens entre les categories d'age et l'usage des reseaux sociaux pour ce faire nous allons utiliser la base de donnée de l'ARCOM. Dans ce contexte, l'analyse factorielle des correspondances s'inscrit pleinement dans la démarche adoptée. Elle permet d'étudier de manière synthétique et visuelle la relation entre deux variables qualitatives : les classes d'âge et les réseaux sociaux utilisés comme sources d'information.

Cette analyse constitue une première étape dans l'étude des pratiques d'information du public. Elle permet d'identifier des logiques générationnelles dans l'usage des réseaux sociaux, qui seront ensuite approfondies dans la Partie III à l'aide d'une Analyse des Correspondances Multiples, intégrant un ensemble plus large de variables qualitatives relatives aux perceptions et comportements informationnels.

L'analyse factorielle des correspondances (AFC) est utilisée afin d'étudier la relation entre les classes d'âge et les réseaux sociaux utilisés comme sources d'information. Avant toute interprétation, un test du khi-deux d'indépendance est réalisé.

Table 1: Résultats du test du khi-deux

	Statistique_chi2	Degres_de_liberte	p.value
X-squared	679.14	48	2.3e-112

Les résultats du test montrent une statistique de $\chi^2 = 679.14$ pour 48 degrés de liberté, avec une p-value inférieure à $2,3 \times 10^{-112}$. L'hypothèse d'indépendance est donc rejetée, ce qui confirme l'existence d'un lien statistiquement significatif entre l'âge et l'usage des réseaux sociaux, justifiant le recours à l'AFC.

4 Partie III – ACM

4.1

4.2

5 Partie IV – Clustering

5.1

5.2

6 Conclusion générale

7 Annexe