

Analyse de données sur l'ISF en France.





Sommaire :

I) Introduction.....	
1. Contexte du choix de la base de données.....	
2. Objectif de l'analyse.....	
II) Méthodologie, constitution de la base de données.....	
1. Choix de la source de données.....	
2. Traitement et organisation des données.....	
III) Méthodologie de l'analyse.....	
1. Analyse factorielle des correspondances.....	
2. Analyse des individus et variables actifs.....	
IV) Synthèse et conclusion.....	
V) Tables des Annexes.....	



I) Introduction :

L'impôt de solidarité sur la fortune (ISF) est l'ancien impôt sur la fortune français payé par les personnes physiques et les couples détenant un patrimoine net taxable strictement supérieur à un certain seuil d'entrée au 1er janvier de l'année considérée. L'ISF est souvent considéré comme une « exception française », vu son absence dans la plupart des autres pays du monde. Souvent qualifié « d'impôt idéologique », il est critiqué pour des raisons morales, fiscales ou économiques. Tous les biens du foyer fiscal sont pris en compte (biens immobiliers, patrimoine financier, véhicules, biens professionnels ne remplissant pas les conditions d'exonération. Ici, dans le cadre de l'ISF, c'est le déclarant, donc le contribuable, qui évalue lui-même la valeur de ses biens.

1) Contexte du choix de la base de données :

Nous avons alors décidé de nous pencher sur le sujet et de chercher une base de données en rapport avec l'ISF en France. Après plusieurs recherches sur le site “data.gouv” regroupant plusieurs base de données, nous avons décidé d'étudier un jeu de données qui pour chaque commune de plus de vingt-mille habitants ayant plus de cinquante redevables à l'Impôt de solidarité sur la fortune (ISF), nous permet d'avoir les informations suivantes : le nombre de redevables, le patrimoine moyen et la cotisation moyenne.

2) Objectif de l'analyse :

Le but de notre projet est de déterminer le rapport entre le nombre de redevables, leurs patrimoines ainsi que le prix au mètre carré du lieu d'habitation. Nous cherchons à comprendre les conditions d'un citoyen Français payant l'ISF, paye-t-il l'ISF dans une commune avec un prix au mètre carré élevé ou plutôt faible, un patrimoine élevé ou faible, nous comparerons également quelles sont les communes où il y a le plus de redevables. Nous avons décidé par la suite d'utiliser une AFC, car c'est la méthode la plus pertinente dans notre cas. Nous avons par la suite sélectionné les variables répondant au thème de notre étude. Ainsi, nous avons dans un premier temps retenu onze variables actives qui nous semblaient être les plus pertinentes pour notre thème. Quels sont les régions comportant le plus de redevables à l'ISF et quel est le rapport avec le prix du logement en fonction de la localisation des redevables ? Nous allons répondre à cette problématique à partir d'une analyse en AFC.



II) Méthodologie, constitution de la base de données

1) Choix de la source de données :

Dans un premier temps nous avons travaillé avec les données disponibles de 2015 à 2017, mais le nombre de variables étant limités, avec le conseil de notre enseignant, nous avons décidé de rajouter le prix moyen au mètre carré de chaque région ainsi que le nombre d'habitants dans chaque département que nous sommes allés chercher sur internet.

2) Traitement et organisation des données :

Après l'ajout de nouvelles variables, nous avons réorganisé la base sur Excel, nous avons choisi onze variables actives et quatre variables supplémentaires qui nous seront utiles à l'interprétation des résultats. Nous avons également supprimé les résultats de l'année 2014 et 2003, énormément de villes manquaient dans les fichiers Excel fournis, nous avons donc décidé de ne pas les utiliser pour ne pas fausser les résultats.

Pour finir, nous avons effectué l'analyse de données sur le logiciel R. L'intégralité des codes R sont disponibles dans la partie Annexes.

Notre tableau comporte onze variables qualitatives actives :

- Population selon le département
- Prix au mètre carré moyen en euro de chaque département
- Nombre de redevables dans chaque commune (2015,2016,2017)
- Patrimoine moyen en euro d'un individu dans une commune (2015,2016,2017)
- Impôt moyen en euro d'un individu dans une commune (2015,2016,2017)

Nous avons quatre variables qualitatives supplémentaires :

- Région
- Département
- Code commune
- Commune



III) Méthodologie de l'analyse

1) Analyse factorielle des correspondances :

1.1) Choix de la méthode :

Définition : L'analyse factorielle des correspondances (AFC) est une technique de visualisation très populaire en analyse de données. Elle est adaptée à l'étude des tableaux de contingence (deux variables qualitatives). Un des objectifs est de produire une représentation, dans un repère unique, des catégories en lignes et en colonnes du tableau afin de mettre en évidence leurs positions respectives, et les éventuelles attractions-répulsions entre les caractéristiques.

Dans notre cas nous avons bien un tableau de contingences entre deux variables qualitatives avec plusieurs cellules, d'où le choix de la méthode en AFC.

1.2) Sélection des valeurs propres :

Afin d'interpréter cette AFC, il faut d'abord déterminer le nombre d'axes que nous allons étudier. Pour cela il existe deux méthodes. Le critère du coude ou le critère de Kaiser.

Avec la règle de Kaiser, nous retiendrons que les axes dont l'inertie est supérieure à l'inertie moyenne, on obtient qu'une seule valeur propre, cela n'est pas très adapté pour notre étude comme on peut le voir dans le tableau.

Tableau 1 : Tableau des valeurs propres (critère de Kaiser)

Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10
2.9278 06e-02	3.215 657e-03	5.0461 21e-04	3.0167 09e-04	8.0234 58e-05	6.3627 83e-05	4.0573 70e-05	8.7286 66e-06	4.382 582e-06	2.1124 84e-07
Vraie	Faux	Faux	Faux	Faux	Faux	Faux	Faux	Faux	Faux

On décide alors de choisir nos valeurs propres avec le critère du coude, on retient nos deux premiers axes comme on peut le voir sur le graphique suivant.

Graphique 1 : Histogramme des valeurs propres (critère du coude)

Histogramme des valeurs propres

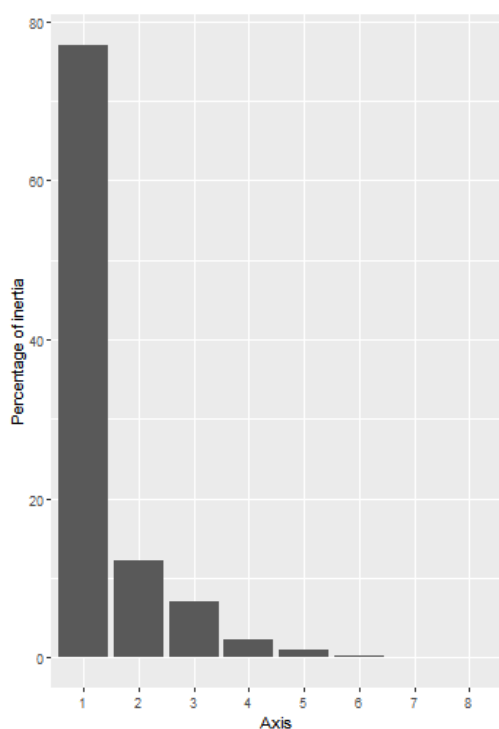


Tableau des valeurs
propres

Axis	%	Cum. %
1	77.0	77.0
2	12.3	89.3
3	7.1	96.4
4	2.3	98.7
5	1.0	99.7
6	0.2	99.9
7	0.1	100.0
8	0.0	100.0

Le premier axe traduit 77% de l'inertie du nuage. La chute est très importante dès le deuxième axe qui ne conserve plus que 12.3% de l'inertie totale. Le troisième axe diminue encore jusqu'à 7.1%. Nous décidons donc de ne retenir que les deux premiers axes. Le troisième axe pourrait être utiliser mais nous jugeons son apport beaucoup trop faible a notre analyse.

2) Analyse des individus et variables actifs :

Graphique 2 : Graphique de corrélation entre les variables



À première vue on remarque que les nombres de redevables en 2015 et en 2017 sont fortement corrélés entre elles. Elles sont également opposées au nombre de redevables en 2016. Le reste des variables sont fortement concentrées au centre du graphique, nous allons étudier plus précisément ces variables. (Petite coquille dans le tableau Excel, une case vide a décalé le nombre de redevables en 2016 d'une case, cela a été corrigé dans la suite pour l'interprétation)

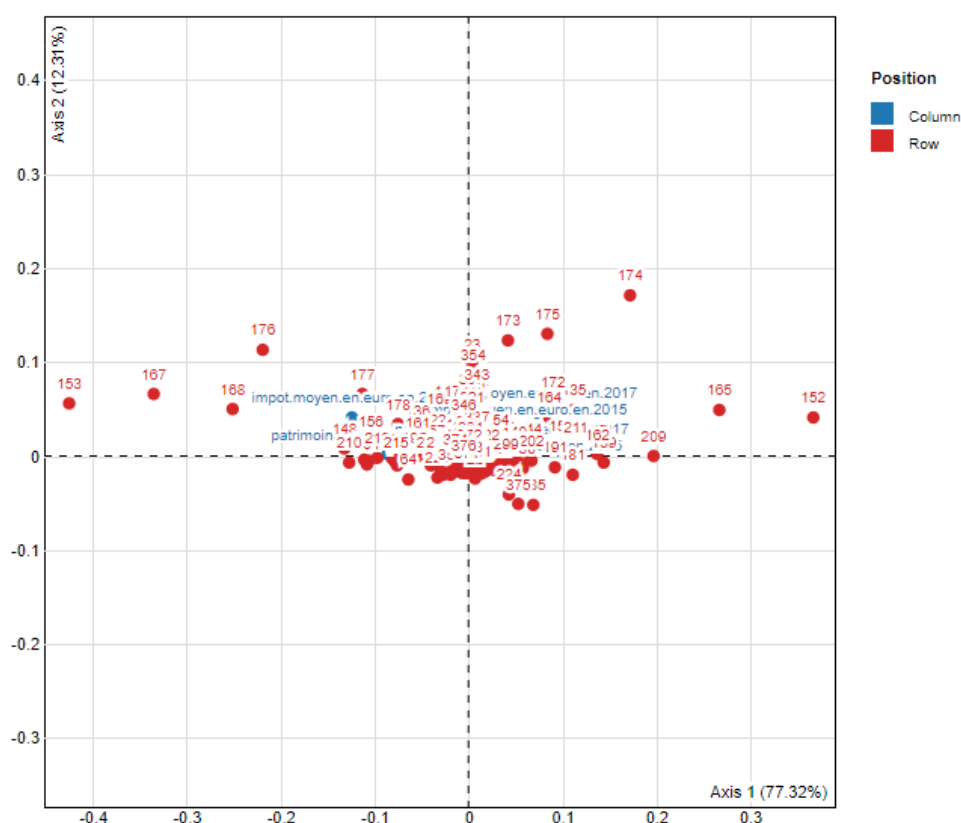
Tableau 2 : Corrélation entre les variables actives selon les axes (colonnes)

	Axe 1	Axe 2
Redevable 2015	0.612	1.720
Redevable 2016	-0.632	1.558
Redevable 2017	0.586	1.699
Patrimoine 2015	0.039	-0.010
Patrimoine 2016	-0.086	0.002
Patrimoine 2017	0.046	0.007
Impôt 2015	0.062	0.029
Impôt 2016	-0.124	0.041
Impôt 2017	0.073	0.046

D'après le tableau 2, sur l'axe 1, l'année 2016 dans les 3 catégories de variables est négativement corrélée aux années 2015 et 2017 qui sont positivement corrélées entre elles.

Sur l'axe 2, on observe un groupe de variables composées du nombre de redevables en 2015, 2016 et 2017 qui sont opposé au patrimoine et aux impôts. Le patrimoine et les impôts sont fortement corrélés entre eux sur l'axe 2.

Graphique 3 : Graphique de corrélation entre les individus



Pour pouvoir analyser ce nuage de points en AFC, il est préférable de faire les comparaisons entre lignes et colonnes.

On observe sur le graphique ci-dessus, que les individus numéro 152 et 165 sont corrélés positivement aux individus 153 et 167 corrélés négativement. D'après notre base de données, la commune de Nanterre compte très peu de redevables tandis que la commune de Neuilly-Sur-Seine compte énormément de redevables.



Selon l'axe 1 nous allons avoir une opposition entre les communes avec un grand nombre de redevable et les communes avec très peu de redevables, tandis que l'axe 2 oppose le nombre de redevables au patrimoine moyen. On remarque que la commune de Paris contient énormément de redevable avec un prix du logement au mètre carré plutôt élevé qui est opposé aux communes de la Réunion ou encore de la Gironde qui ont un prix au mètre carré plutôt faible avec très peu de redevable. Les redevables qui payent l'ISF sont plutôt concentré dans les communes avec un grand nombre d'habitants, Paris, Marseille ou encore Lyon, là où le logement a un prix au mètre carré très important. Le patrimoine moyen d'un habitant de Paris est très différent d'un habitant de la Réunion, cela justifie le prix de l'immobilier en Ile de la Réunion.

On remarque par exemple une opposition entre les individus 05 et 200 selon l'axe 2. Ce qui correspond à Vichy et Rosny sous-bois, Vichy a un faible prix au mètre carré : 1028 euros contrairement à Rosny sous-bois avec 3813 euros. Globalement les communes avec un fort prix au mètre carré paient un impôt moyen plus élevé.

IV) Synthèse et conclusion

Pour conclure on dira que cette AFC nous a permis de constater une certaine corrélation entre l'impôt et le patrimoine comme nous l'avait indiqué le graphique de corrélation entre les variables. On a globalement observé que les communes où de nombreux habitants payaient l'ISF étaient caractérisées par une forte population ainsi qu'un prix au mètre carré élevé et un impôt moyen plus élevé.



V) Tables des annexes

Code sous le logiciel R :

```
# Lecture des packages nécessaires

library(FactoMineR)

library(explor)

# Répertoire du fichier

setwd("C:/Users/Wass/Desktop/Analyse de donnée/Projet")

# Lecture du fichier

projet<-read.csv(file="baseprojet.csv")

#Connaitre les dimensions de la base de données

dim(projet)

#On supprime les variables qualitatives

quali.sup=1:6

#Matrice de corrélation

cor(projet[-1,7:15])

#On effectue une AFC sur la matrice des corrélations

res.afc<-CA(projet[-1,7:15])

#On affiche dans explore

explor(res.afc)

-----

#On calcul les VP ainsi que leurs sommes

vp<-res.afc$eig

vp[,1]

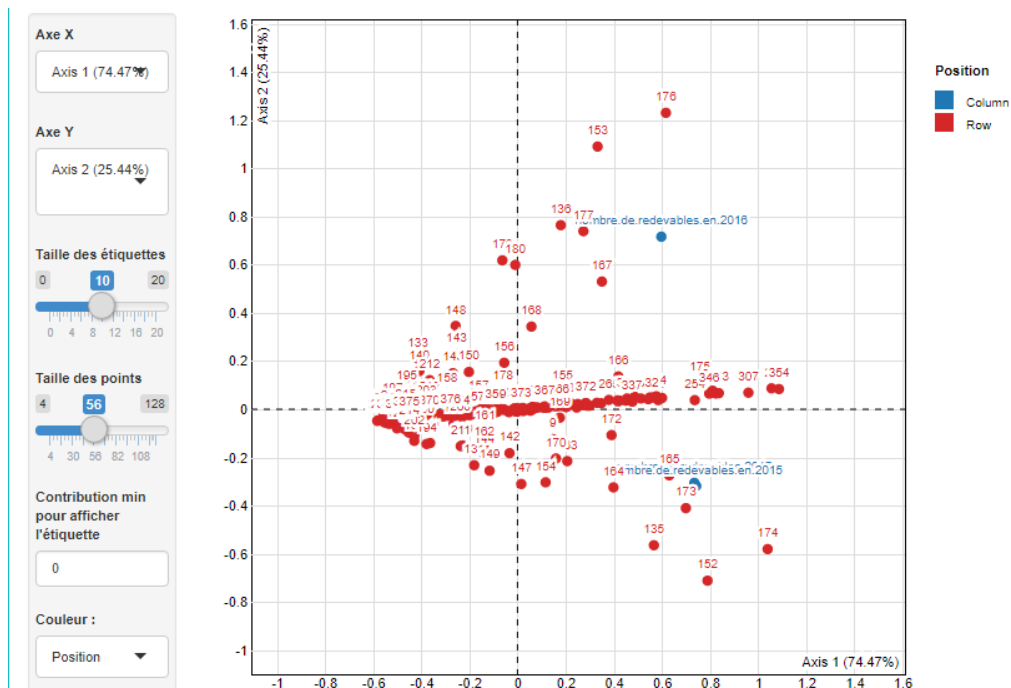
#On conserve les valeurs propres supérieur a la moyenne

vp[,1]>mean(vp[,1])

#On affiche de nouveau dans explor

explor(res.afc)
```

Graphique 3 : Prix au mètre carré selon les redevables des communes



Graphique 4 : Impôt moyen selon les communes

