

Analyse spatiale et territoriale de données de recensement

Formation Carthageo-Geoprisme 2021 / 1ere journée

Claude Grasland (U. Paris Diderot)

04/10/2021

Section 1

1.ANALYSE SOCIALE : RELATION ENTRE LOGEMENT HLM ET CSP

Définir le sujet

Soit le sujet : *Logements sociaux et classes sociales dans le Val de Marne en 2017*

Définir les “logements sociaux”

Logements HLM ? Logements SRU ?

Définir la notion de “classes sociales” ?

Les CSP ? Un regroupement de CSP ? Les revenus ? la génération

Définir la date

Année 2017 uniquement ? Résultats du RP 2017 (2015-2019) ?

Formuler des questions ou des hypothèses

Qu'elles soient justes ou fausses, les hypothèses permettent de cadrer l'analyse.

CSP et logement social

Les logements sociaux sont réservés aux CSP les plus modestes

Âge et logement social

Les logements sociaux sont réservés aux jeunes ménages actifs

Logement social et territoire

Les logements sociaux sont concentrés dans certains quartiers

Logement social, âge et territoires

Les personnes à la retraite quittent les logements sociaux.

Organiser le travail

Surtout dans le cadre d'un groupe !

Ne collecter que les données utiles pour répondre aux questions posées

Afin de ne pas être tenté de partir dans toutes les directions

Archiver soigneusement les programmes et les résultats

Afin de pouvoir reproduire ultérieurement les analyses sur une autre période, un autre territoire

Ne pas attendre d'avoir accumulé tous les résultats pour les commenter

Car l'analyse peut suggérer des erreurs ou ouvrir de nouvelles pistes.

Partir des questions et non pas des outils

Faute de quoi on va trouver des réponses (42 ...) sans savoir quelle est la question.

Ne pas confondre les niveaux d'agrégation

Les réponses peuvent varier selon le niveau d'agrégation.

individu / groupe social

“Un ouvrier a plus de chance d'habiter en HLM qu'un cadre”

individu / groupe territorial

“Les cadres et les ouvriers n'habitent pas dans les mêmes quartiers”

groupe social / groupe territorial

“Les cadres évitent les quartiers où il y a beaucoup de logements sociaux”

Charger les données statistiques

programme

```
tab_ind<-readRDS("data2021/94/indiv2017.RDS")
```

résultat

FALSE	CANTVILLE	NUMMI	ACHLR	AEMMR	AGED
FALSE 1	9401	1	3	9	47
FALSE 2	9401	1	3	9	42

Préparation de l'analyse

- Soit la relation entre logement en HLM (Y) et CSP du chef de ménage (X). Il s'agit de deux variables **catégorielles** (= qualitatives) que l'on va typiquement mettre en relation à l'aide d'un **tableau de contingence** et d'un **test du chi-2**. L'analyse statistique est simple sous R mais il faut tenir compte de trois difficultés
- Le choix de la **population de référence** est important
- la sélection ou le regroupement des **CSP** est également important car il va influencer sur les résultats du test.
- la **pondération des individus** doit également être prise en compte puisque le recensement est basé sur un sondage

Sélection des individus et des variables

Puisqu'on travaille sur les logements, on ne garde que les personnes de références des ménages (LPRM=1).

programme

```
#table(tab_ind$AGEMEN8)  
tab_sel<- tab_ind %>%  
  filter(LPRM == 1) %>%  
  select(CS1,HLML, IPONDI)
```

résultats

CS1	HLML	IPONDI
3	1	2.9845995
4	2	1.1925306
8	2	0.8925770
2	2	0.8984924

Recodage des modalités

On cherche le code des modalités CS1 ezt HLML dans le fichier des métadonnées

COD_VAR	COD_MOD	LIB_MOD
CS1	1	Agriculteurs exploitants
CS1	2	Artisans, commerçants et chefs d'entreprise
CS1	3	Cadres et professions intellectuelles supérieures
CS1	4	Professions Intermédiaires
CS1	5	Employés
CS1	6	Ouvriers
CS1	7	Retraités
CS1	8	Autres personnes sans activité professionnelle
HLML	1	Logement appartenant à un organisme HLM
HLML	2	Logement n'appartenant pas à un organisme HLM
HLML	Z	Hors logement ordinaire

Recodage des modalités

On recode les modalités des deux variables en regroupant certaines CSP

programme

```
tab_sel$HLML<-as.factor(tab_sel$HLML)
levels(tab_sel$HLML)<-c("HLM-O", "HLM-N")
tab_sel$CS1<-as.factor(tab_sel$CS1)
levels(tab_sel$CS1) <- c("ARCAD", "ARCAD", "ARCAD", "INTER",
                        "EMPOU", "EMPOU", "RETRA", "INACT")
```

résultats

CS1	HLML	IPONDI
ARCAD	HLM-O	2.984599
INTER	HLM-N	1.192531
INACT	HLM-N	0.892577

Création du tableau de contingence non pondéré (FAUX)

La solution la plus simple semble être l'instruction `table()`

programme

```
tab_cont<-table(tab_sel$HLML,tab_sel$CS1)
```

résultats

	ARCAD	INTER	EMPOU	RETRA	INACT	Sum
HLM-O	6246	11785	32802	12824	4728	68385
HLM-N	46158	30804	38272	34803	9865	159902
Sum	52404	42589	71074	47627	14593	228287

Création du tableau de contingence pondéré (JUSTE)

On pondère avec `wtd.table()` du package *questionr*.

programme

```
library(questionr)
tab_cont_wtd<-wtd.table(tab_sel$HML,tab_sel$CS1,
                        weights = tab_sel$IPONDI)
```

résultats

	ARCAD	INTER	EMPOU	RETRA	INACT	Sum
HLM-O	15193	28327	77954	31314	10381	163169
HLM-N	123760	80991	99145	98388	21225	423509
Sum	138952	109318	177099	129702	31607	586678

Comparaison des niveaux de dépendance automobile

- Tableau non pondéré ... légèrement faux !

	ARCAD	INTER	EMPOU	RETRA	INACT	Ensemble
HLM-O	11.9	27.7	46.2	26.9	32.4	30
HLM-N	88.1	72.3	53.8	73.1	67.6	70
Total	100.0	100.0	100.0	100.0	100.0	100

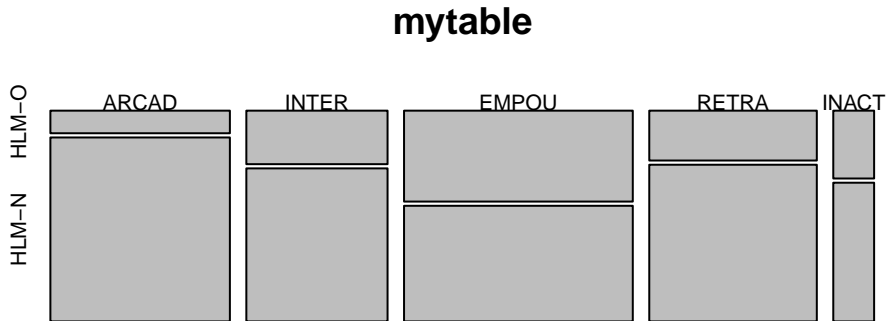
- Tableau pondéré ... juste !

	ARCAD	INTER	EMPOU	RETRA	INACT	Ensemble
HLM-O	10.9	25.9	44	24.1	32.8	27.8
HLM-N	89.1	74.1	56	75.9	67.2	72.2
Total	100.0	100.0	100	100.0	100.0	100.0

Visualisation du tableau de contingence

On choisit l'orientation du tableau et on l'affiche avec `plot()`

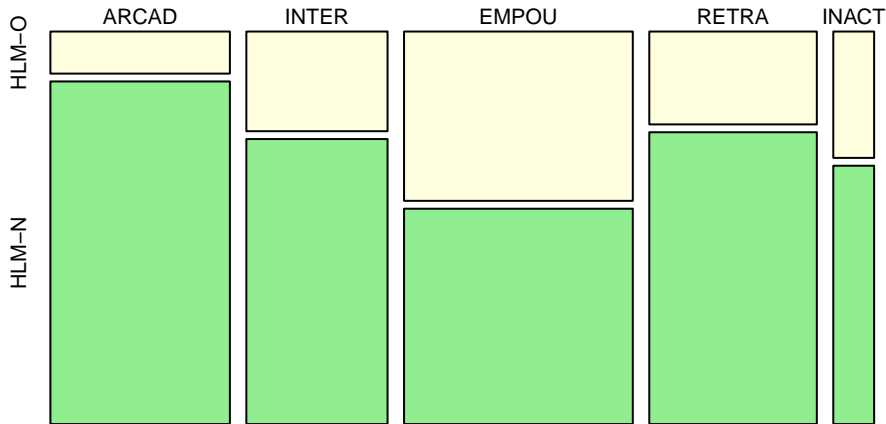
```
mytable<-wtd.table(tab_sel$CS1,tab_sel$HLMML,weights = tab_sel$  
plot(mytable)
```



Visualisation améliorée du tableau de contingence

Tant qu'à faire, on améliore la figure avec des paramètres supplémentaires :

Logements HLM par CSP



Test du Chi-deux

Ce test se réalise facilement sur le tableau de contingence avec l'instruction *chisq.test()* :

```
mytest<-chisq.test(mytable)
mytest
```

Pearson's Chi-squared test

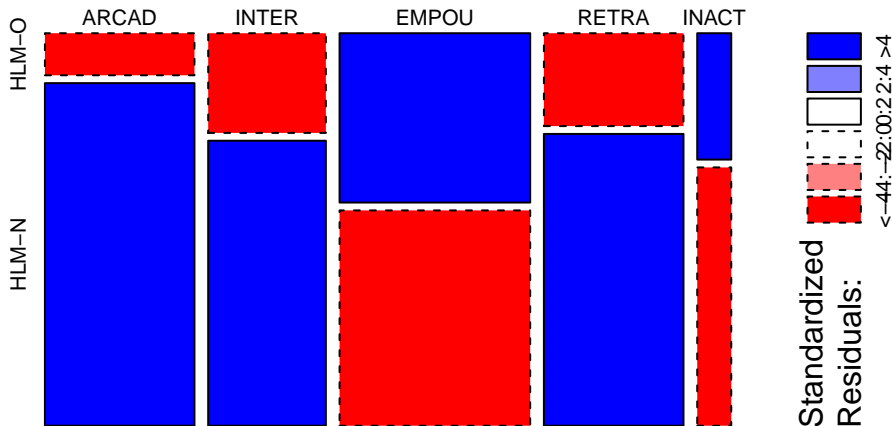
data: mytable

X-squared = 44346, df = 4, p-value < 2.2e-16

Visualisation des résidus

Lorsque la relation est significative, on visualise les cases les plus exceptionnelles avec `mosaicplot(..., shade = T)`

Logements HLM par CSP



Conclusion

28% des ménages ordinaires du Val de Marne résident en HLM

Ce chiffre ne tient toutefois pas compte de la **population concernée** qui peut être plus élevée.

La part du logement en HLM varie bien en fonction de la CSP des actifs

44% des employés et ouvriers, 26% des intermédiaires, 11% des cadres.

Les inactifs sont également nombreux en HLM

33% d'entre eux sont dans ce cas, recouvrant des situations diverses.

Une partie des retraités demeurent en HLM

Environ 25% d'entre eux.