

Analyse spatiale et territoriale du logement social

Formation Carthageo-Geoprisme 2022

Claude Grasland, Université de Paris (Diderot)

Section 1

Données RP 2018

Définir le sujet

Soit le sujet : *Logements sociaux et qualification des chefs de ménages*

Définir les “logements sociaux”

Logements HLM ? Logements SRU ?

Définir la notion de “qualification” ?

Le diplôme le plus élevé ? le nombre d'années d'étude ?

Définir la date

Année 2018 uniquement ? Résultats du RP 2018 (2016-2020) ?

Formuler des questions ou des hypothèses

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

Diplôme et logement social

Les logements sociaux sont réservés aux ménages les moins diplômés

Âge et logement social

Les logements sociaux sont réservés aux jeunes ménages

Logement social et territoire

Les logements sociaux sont concentrés dans certains quartiers

Logement social, âge et diplômes

Les personnes diplômées quittent les logements sociaux dès que leurs revenus progressent

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.

Charger les données statistiques

programme

```
tab_ind<-readRDS("data/menag2018.RDS")
```

résultat

FALSE	COMMUNE	ARM	IRIS	ACHL	AEMM
FALSE 1	75056	75101	751010101	C114	2014
FALSE 2	75056	75101	751010101	A11	2008

Préparation de l'analyse

- Soit la relation entre logement en HLM (Y) et Diplôme le plus élevé 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. Ici on va sélectionner les ménages dont la personne de référence est âgée de 25-39 ans
- la sélection ou le regroupement des **diplômes** est également important car cela 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

programme

```
#table(tab_ind$AGEMEN8)
tab_sel<- tab_ind %>%
  filter(AGEMEN8 == "25") %>%
  select(DIPLM, HLML, IPONDL)
```

résultats

DIPLM	HLML	IPONDL
18	2	2.628217
18	2	2.628217
18	2	2.639089
18	2	2.574046

Recodage des modalités

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

COD_VOC	COD_MOD	MOD
---------	---------	-----

DIPLM01	Pas de scolarité ou arrêt avant la fin du primaire
DIPLM02	Aucun diplôme et scolarité interrompue à la fin du primaire ou avant la fin du collège
DIPLM03	Aucun diplôme et scolarité jusqu'à la fin du collège ou au-delà
DIPLM11	CEP (certificat d'études primaires)
DIPLM12	BEPC, brevet élémentaire, brevet des collèges, DNB
DIPLM13	CAP, BEP ou diplôme de niveau équivalent
DIPLM14	Baccalauréat général ou technologique, brevet supérieur, capacité en droit, DAEU, ESEU
DIPLM15	Baccalauréat professionnel, brevet professionnel, de technicien ou d'enseignement, diplôme équivalent
DIPLM16	BTS, DUT, Deug, Deust, diplôme de la santé ou du social

Recodage des modalités

On recode les modalités des deux variables en regroupant certaines CSP

programme

```
tab_sel$HMLM<-as.factor(tab_sel$HMLM)
levels(tab_sel$HMLM)<-c("HLM-O", "HLM-N", NA)
tab_sel$DIPLM<-as.factor(tab_sel$DIPLM)
levels(tab_sel$DIPLM) <- c("< BAC", "< BAC", "< BAC", "< BAC", "< BAC",
                           "BAC", "BAC",
                           "BAC+123", "BAC+123", "> BAC+3", "> BAC+3")
table(tab_sel$DIPLM)
```

FALSE

FALSE < BAC BAC BAC+123 > BAC+3

FALSE 65376 50352 88531 156494

résultats

Création du tableau de contingence non pondéré (FAUX)

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

programme

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

résultats

	< BAC	BAC	BAC+123	> BAC+3	Sum
HLM-O	27517	16957	18316	12692	75482
HLM-N	37859	33395	70215	143802	285271
Sum	65376	50352	88531	156494	360753

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$HLML,tab_sel$DIPLM,
                        weights = tab_sel$IPONDL)
```

résultats

	< BAC	BAC	BAC+123	> BAC+3	Sum
HLM-O	62959	38267	41036	27175	169437
HLM-N	88580	77395	166707	356126	688808
Sum	151539	115662	207742	383302	858245

Comparaison des niveaux de dépendance automobile

- Tableau non pondéré ... légèrement faux !

	< BAC	BAC	BAC+123	> BAC+3	Ensemble
HLM-O	42.1	33.7	20.7	8.1	20.9
HLM-N	57.9	66.3	79.3	91.9	79.1
Total	100.0	100.0	100.0	100.0	100.0

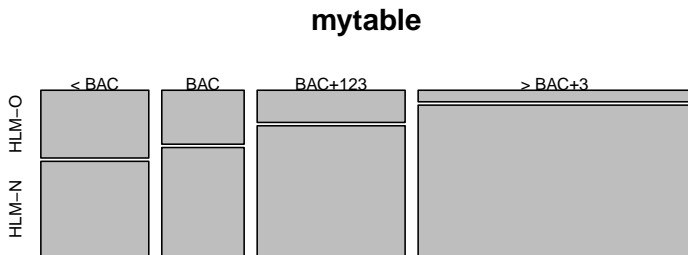
- Tableau pondéré ... juste !

	< BAC	BAC	BAC+123	> BAC+3	Ensemble
HLM-O	41.5	33.1	19.8	7.1	19.7
HLM-N	58.5	66.9	80.2	92.9	80.3
Total	100.0	100.0	100.0	100.0	100.0

Visualisation du tableau de contingence

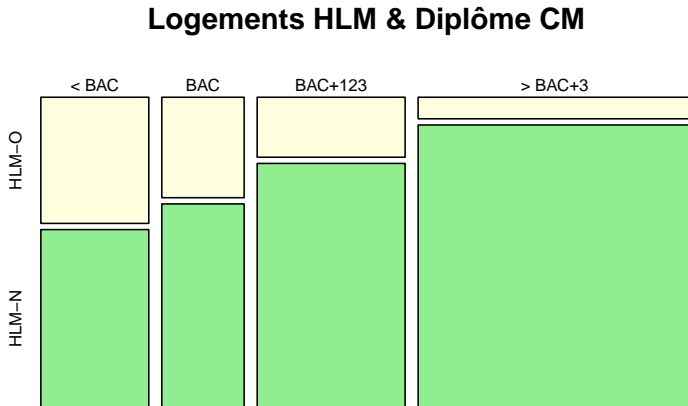
On choisit l'orientation du tableau et on l'affiche avec `plot()`

```
mytable<-wtd.table(tab_sel$DIPLM,tab_sel$HLML,weights = tab_sel$weights)
plot(mytable)
```



Visualisation améliorée du tableau de contingence

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



Source : INSEE – RP 2018

Test du Chi-deux

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

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

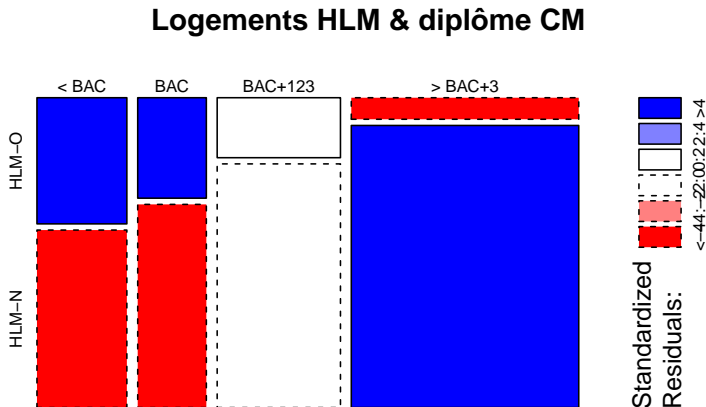
Pearson's Chi-squared test

data: mytable

X-squared = 97192, df = 3, p-value < 2.2e-16

Visualisation des résidus

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



Source : INSEE – RP 2018

Conclusion