

Analyse en Composantes Principales (ACP)

Quentin Garnier

11 mars 2025

```
library(FactoMineR)
```

```
## Warning: le package 'FactoMineR' a été compilé avec la version R 4.3.3
```

```
credit0 <- read.csv(file("./credit.csv"),header=TRUE,sep=";",row.names=1)
credit <- credit0[-68,c(1,2,3,10,11,4,5,6,7,8,9)]
summary(credit)
```

```
##      marché      Apport      Impayé      assurance
## Length:67      Length:67      Length:67      Length:67
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
## endettement      Famille      Enfants      Logement
## Length:67      Length:67      Length:67      Length:67
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
## Profession      Intitulé      Age
## Length:67      Length:67      Length:67
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
```

```
for (i in 1:ncol(credit)) credit[,i] <- factor(as.character(credit[,i]))
attach(credit)
summary(credit)
```

```
##      marché      Apport      Impayé      assurance
## ile      : 1      Non:32      0      :43      AID      :31
## Mobilier / Ameublement:17      Oui:35      1 ou 2: 8      AID + Chomage:13
## Moto      : 9      3 et +:16      Sans      :13
## Rénovation      :18      Senior      :10
## Scooter      : 5
## Voiture      :17
## endettement      Famille      Enfants      Logement
## 1:19      Célibataire:17      0:39      Accédant à la propriété: 6
## 2:15      Divorcé      : 5      1: 8      Locataire      :23
## 3:19      Marié      :25      2:11      Logé par l'employeur : 3
## 4:14      Pacsé      : 1      3: 6      Logé par la famille : 6
##      Union libre:13      4: 2      Propriétaire      :29
##      Veuf      : 6      8: 1
```

```
##          Profession Intitulé Age
## Cadre moyen      :17  MLLE: 5  20:10
## Cadre sup.       : 8   MME : 9  30:21
## Ministre         : 1   MR  :53  40:10
## Ouvrier non qualifié:11          50:14
## Ouvrier qualifié  :18          60:12
## Retraité         :12
```

3. Que faut il avant de commencer une AFCM ?

3.1. Vérification des données

- **Variables catégorielles** uniquement (nominales ou ordinales).
- Format : individus en ligne, variables en colonne.

3.2. Taille de l'échantillon

- ≥ 5 à 10 fois le nombre total de modalités.

3.3. Modalités rares

- Éviter celles avec < 5 occurrences.
- Fusionner les modalités proches si nécessaire.

3.4. Relations entre variables

- Vérifier la **corrélation** et **redondance**.

3.5. Valeurs manquantes

- Supprimer ou remplacer par une modalité "Non Répondu".

5. Analyse des valeurs propres et du nombre d'axes retenus

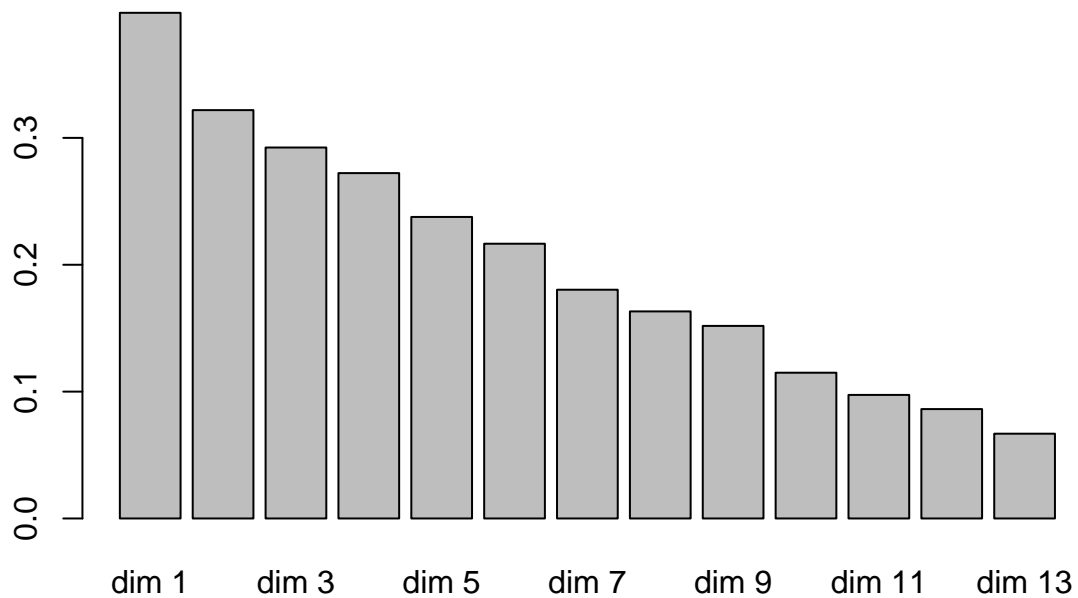
```
help(MCA)
# res.MCA=MCA(credit,quali.sup=6:11)
res.MCA <- MCA(credit[-67,],quali.sup=6:11,graph=FALSE)

res.MCA$eig
```

```
##          eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.39858654          15.330251          15.33025
## dim 2  0.32188055          12.380021          27.71027
## dim 3  0.29241671          11.246797          38.95707
## dim 4  0.27223718          10.470661          49.42773
## dim 5  0.23766789           9.141073          58.56880
```

## dim 6	0.21655909	8.329196	66.89800
## dim 7	0.18031383	6.935147	73.83315
## dim 8	0.16322576	6.277914	80.11106
## dim 9	0.15178184	5.837763	85.94882
## dim 10	0.11489708	4.419118	90.36794
## dim 11	0.09739129	3.745819	94.11376
## dim 12	0.08622601	3.316385	97.43015
## dim 13	0.06681622	2.569855	100.00000

```
barplot(res.MCA$eig[,1])
```



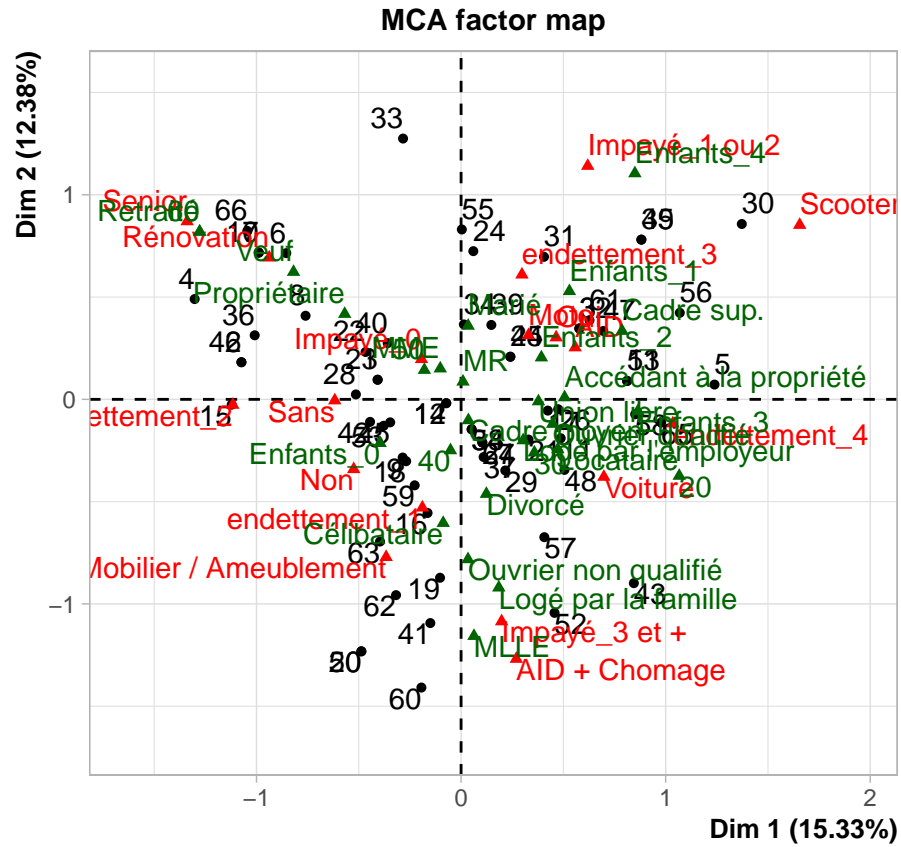
Les valeurs propres indiquent la variance expliquée par chaque axe :

- Le premier axe capture **15.33%** de la variance totale.
- Les **4 premiers axes** expliquent **58.58%** de la variance.
- Pour atteindre **85.95%**, il faut **9 axes**, mais les derniers apportent peu d'information.

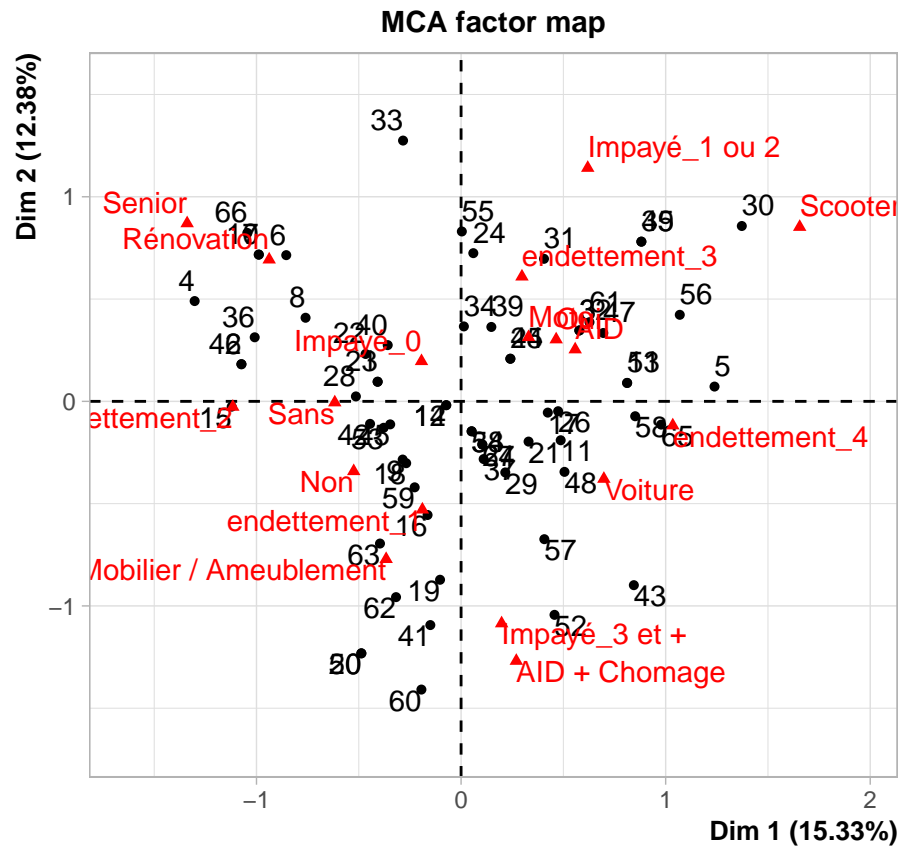
Sélection des axes :

- **Règle du coude** : Forte chute après **4 ou 5 axes**.
- **Interprétation** : Les premiers axes sont les plus pertinents.

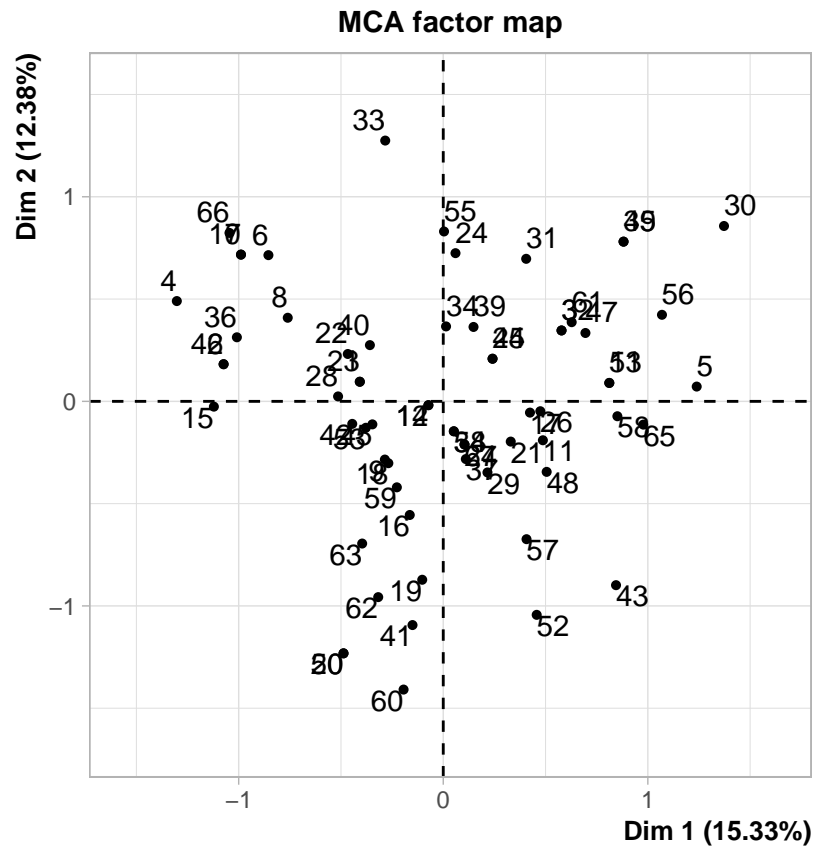
```
help(plot.MCA)
plot.MCA(res.MCA, choix="ind", col.ind="black")
```



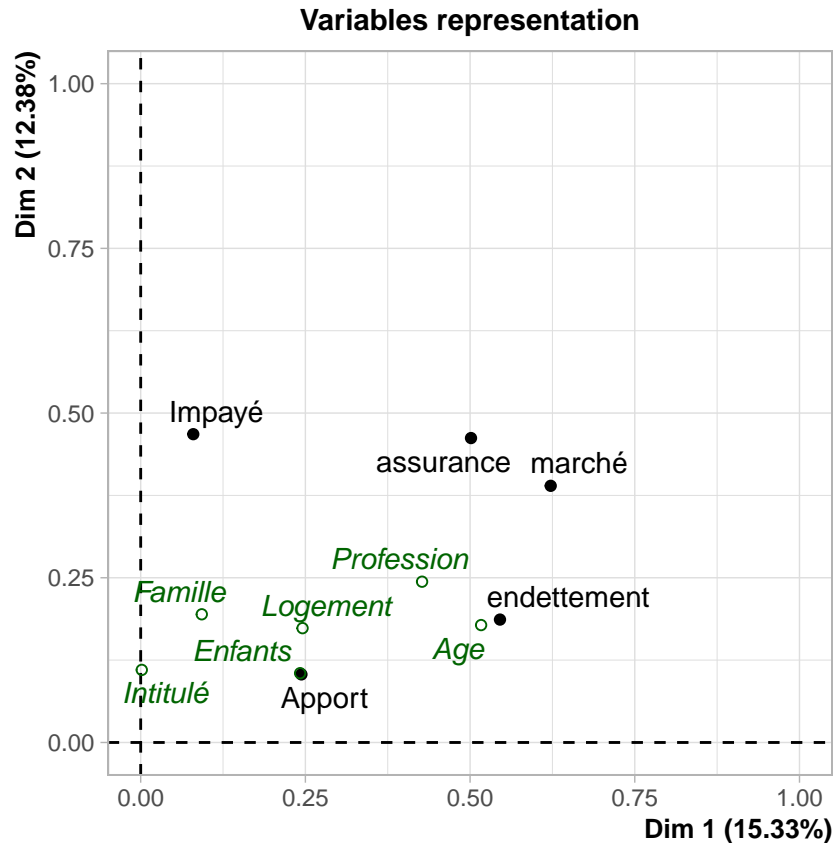
```
plot.MCA(res.MCA, choix="ind", invisible=c("quali.sup", "ind.sup"),
col.ind="black")
```



```
plot.MCA(res.MCA, choix="ind", invisible=c("var", "quali.sup", "ind.sup"),
col.ind="black")
```



```
plot.MCA(res.MCA, choix="var", col.var="black")
```



6. Analyse globale de l'ACM

6.1. Analyse des variables

- Dim 1 (15.33%) : Oppose les profils **financièrement stables vs endettés**.
- Dim 2 (12.38%) : Différencie les **profils socio-économiques** (profession, logement).

6.2. Analyse des individus

- Dispersion importante → bonne différenciation des profils.

6.3. Analyse conjointe

- Groupes identifiables :
- Endettement élevé et impayés (rouge).
- Statut socio-professionnel et logement (vert).
- Corrélations observées : propriétaires vs locataires, endettement élevé vs situation stable.
- Variables corrélées : Enfant et Apport, Age et Endettement.

7. Étude des variables les plus liées aux axes

```
nbvar <- 5
variable <- NULL
for (i in 1:nbvar) {
  variable <- c(variable, rep(names(credit)[i], length(unique(factor(as.character(credit[-67, i]))))))
}

ctr <- res.MCA$var$contrib

# Somme des contributions par variable sur l'axe 1
contrib_axe1 <- tapply(ctr[,1], variable, sum)

# Somme des contributions par variable sur l'axe 2
contrib_axe2 <- tapply(ctr[,2], variable, sum)

# Rapport de corrélation par rapport à l'axe 1
correlation_axe1 <- contrib_axe1 * nbvar * res.MCA$eig[1,1]

# Rapport de corrélation par rapport à l'axe 2
correlation_axe2 <- contrib_axe2 * nbvar * res.MCA$eig[2,1]

list(Axe1 = contrib_axe1, Corr_Axe1 = correlation_axe1,
     Axe2 = contrib_axe2, Corr_Axe2 = correlation_axe2)
```

```
## $Axe1
##      Apport  assurance endettement      Impayé      marché
##  12.241848  25.159214  27.366416    4.003392  31.229129
##
## $Corr_Axe1
##      Apport  assurance endettement      Impayé      marché
##  24.397178  50.140620  54.539426    7.978492  62.237553
##
## $Axe2
##      Apport  assurance endettement      Impayé      marché
##   6.419384  28.709110  11.589904   29.075990  24.205612
##
## $Corr_Axe2
##      Apport  assurance endettement      Impayé      marché
##   10.33137   46.20452   18.65282   46.79498   38.95658
```

7.1. Variables liées à l'axe 1

- Variables influentes : **Marché (31.23%)**, **Endettement (27.37%)**, **Assurance (25.16%)**.
- L'axe 1 oppose **stabilité financière** vs **endettement**.

7.2. Variables liées à l'axe 2

- Variables influentes : **Impayé (29.08%)**, **Assurance (28.71%)**, **Marché (24.21%)**.
- L'axe 2 distingue **gestion du risque** et **couverture financière**.

8. Etude des modalités

Contributions au axes

```
round(res.MCA$var$contrib[rev(order(res.MCA$var$contrib[,1])),1],2)
```

##	endettement_2	Senior	Rénovation
##	14.20	13.63	12.04
##	endettement_4	Scooter	AID
##	11.40	10.41	7.33
##	Non	Voiture	Oui
##	6.49	6.30	5.75
##	Sans	Impayé_1 ou 2 Mobilier / Ameublement	
##	3.48	2.33	1.73
##	endettement_3	Impayé_0	Moto
##	1.28	1.19	0.74
##	AID + Chomage	endettement_1	Impayé_3 et +
##	0.72	0.49	0.48

```
round(res.MCA$var$contrib[rev(order(res.MCA$var$contrib[,2])),2],2)
```

##	AID + Chomage	Impayé_3 et +	Impayé_1 ou 2
##	19.71	17.75	9.80
##	Mobilier / Ameublement	Rénovation	Senior
##	9.52	8.14	7.13
##	endettement_3	endettement_1	Scooter
##	6.65	4.75	3.42
##	Non	Oui	Voiture
##	3.40	3.02	2.30
##	AID	Impayé_0	Moto
##	1.88	1.52	0.83
##	endettement_4	endettement_2	Sans
##	0.18	0.01	0.00

8.1. Modalités contribuant le plus à l'axe 1

- **Principales modalités** : endettement_2 (14.20%), Senior (13.63%), Rénovation (12.04%), Scooter (10.41%), AID (7.33%).
- Ces modalités sont fortement impliquées dans la **différenciation des profils financiers**.

8.2. Modalités contribuant le plus à l'axe 2

- **Principales modalités** : AID + Chômage (19.71%), Impayé_3 et + (17.75%), Impayé_1 ou 2 (9.80%), Mobilier / Ameublement (9.52%), Rénovation (8.14%).
- L'axe 2 distingue les individus selon leur **situation sociale et leur gestion du risque**.

8.3. Position des modalités

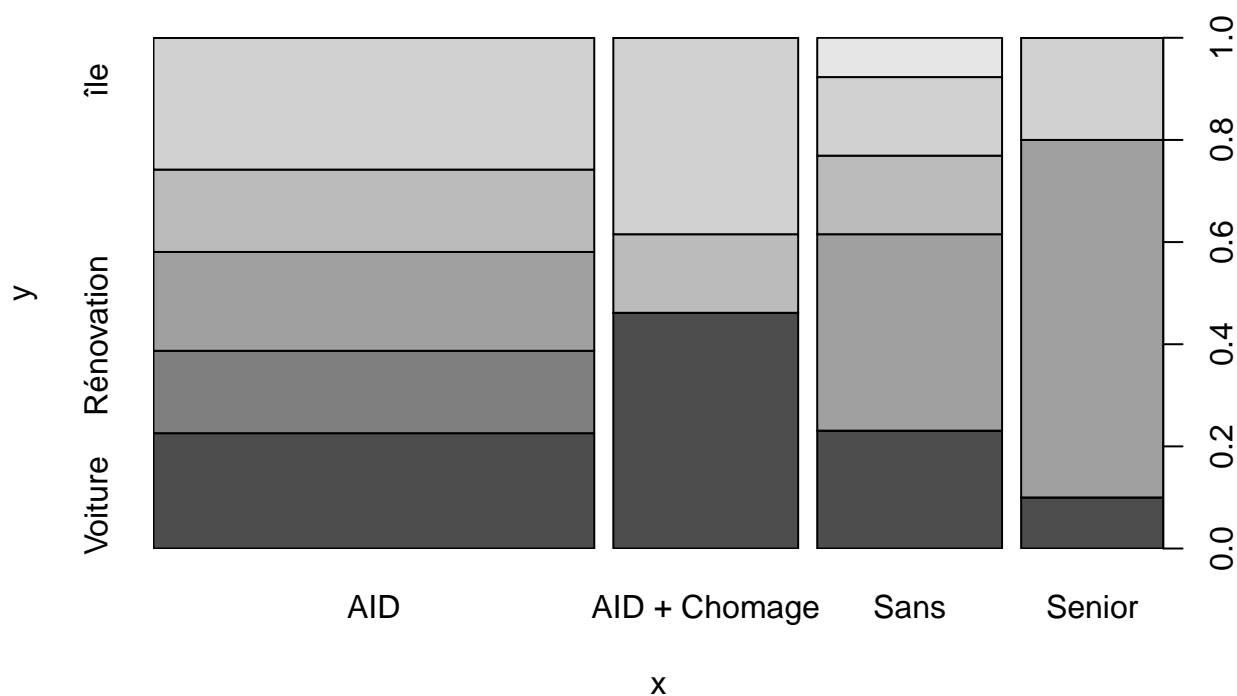
Ces modalités sont **souvent situées aux extrémités** du graphique, ce qui est attendu car elles expliquent la **variabilité maximale**.

8.04. Qualité de représentation et interprétation

- Certaines modalités ont une **forte contribution mais une faible qualité de représentation**.
- Les résultats restent cohérents : les modalités liées à l'endettement, aux impayés et aux aides financières sont les plus discriminantes.

9. Interpretation de la proximité entre Senior et assurance, rénovation et marché

```
plot(assurance,marché)
```



```
conting <- table(assurance,marché)
conting
```

```
##          marché
## assurance  île Mobilier / Ameublement Moto Rénovation Scooter Voiture
##   AID      0          8      5          6      5      7
##   AID + Chomage 0          5      2          0      0      6
##   Sans       1          2      2          5      0      3
##   Senior     0          2      0          7      0      1
```

```
conting <- table(marché, assurance)
conting
```

```
##               assurance
## marché      AID AID + Chomage Sans Senior
##   île                0           0     1     0
##   Mobilier / Ameublement  8           5     2     2
##   Moto                5           2     2     0
##   Rénovation           6           0     5     7
##   Scooter             5           0     0     0
##   Voiture             7           6     3     1
```

9.1. Analyse graphique

- “Senior” (assurance) et “Rénovation” (marché) semblent proches.
- “Rénovation” est plus associée à **AID + Chômage** et **Sans assurance**.

9.2. Vérification avec les données brutes

- **Senior** est fortement lié à **Assurance** (7 occurrences)
- **Renovation** est lié à Senior (7 occurrences) et AID (6 occurrences). ## 9.3. Conclusion
- “Rénovation” est plus liée aux profils précaires (“AID + Chômage”, “Sans assurance”).
- **Les Seniors sont peu impliqués dans les projets de rénovation.**
- Un test du **chi2** permettrait de confirmer ces tendances.