

TP3

AFC et AFCM

Christophe Tiet

Partie I: AFC

1-Justification de l'AFC

```
## Pearson's Chi-squared test
##
## data:  chomtable
## X-squared = 96.57, df = 28, p-value = 1.809e-09
```

Nous trouvons une $p\text{-value} < 0.05$, on rejette donc l'hypothèse nulle d'indépendance et on considère les deux variables comme dépendantes.

Les fortes contributions sont celles > 2.4 ($X^2 = 95.67$ et $n = 40$ donc $X^2/n = 2.4$)

```
test$residuals^2
```

##		A	C	D	I	N
##	1	5.771234919	23.875472810	5.557826663	0.089743590	8.685997739
##	2	0.314181050	2.659129228	0.925944112	3.940428380	0.213704510
##	3	0.109998254	0.019788321	0.085739792	0.480158730	0.570965703
##	4	6.019400289	5.303511893	1.085063393	1.000935194	0.935113290
##	5	0.060945274	0.142018779	0.497787611	0.331453634	2.815315315
##	6	0.060945274	0.620892019	1.382743363	7.173558897	0.004504505
##	7	0.601514604	5.505570405	0.288933024	6.123463000	0.368001434
##	8	0.265731456	1.764619883	0.039181439	0.145891690	0.732663004

```
test$expected
```

##		A	C	D	I	N
##	1	7.640351	8.096491	38.657895	15.166667	8.438596
##	2	16.260234	17.230994	82.271930	32.277778	17.959064
##	3	7.934211	8.407895	40.144737	15.750000	8.763158
##	4	15.378655	16.296784	77.811404	30.527778	16.985380
##	5	5.583333	5.916667	28.250000	11.083333	6.166667
##	6	5.583333	5.916667	28.250000	11.083333	6.166667
##	7	6.954678	7.369883	35.188596	13.805556	7.681287
##	8	1.665205	1.764620	8.425439	3.305556	1.839181

```
test$observed
```

##		A	C	D	I	N
##	1	1	22	24	14	17
##	2	14	24	91	21	16
##	3	7	8	42	13	11
##	4	25	7	87	25	13
##	5	5	5	32	13	2
##	6	5	4	22	20	6
##	7	9	1	32	23	6
##	8	1	0	9	4	3

Les fortes contributions **sous représentées** concernent les couples de variables

1a 1d 2i 4c 5n 7c

Les fortes contributions **sur-représentées** concernent les couples de variables

1c 1n 2c 4a 6i 7i

L'AFC est donc justifiée car nos données sont sous formes de tableau de contingence avec 2 variables qualitatives (anneeetud et emploi). Nous avons également beaucoup de modalités (8 lignes et 5 colonnes) et nos deux variables sont dépendantes.

2-Axes retenus

```
resuCA$eig
```

```
##          eigenvalue percentage of variance cumulative percentage of variance
## dim 1 0.093830096          66.459291          66.45929
## dim 2 0.035482535          25.132066          91.59136
## dim 3 0.008023450           5.682961          97.27432
## dim 4 0.003848235           2.725682         100.00000
```

On décide de retenir deux axes. Le 1er axe a une part d'inertie expliquée de 66.45% tandis que le 2ème axe a une part d'inertie expliquée de 25.13%. En retenant les deux axes, on a une inertie expliquée totale de 91.6%. Et nous ne perdons ainsi que 8.4% de l'information totale.

3-Contribution des profils

```
#Tableau des contributions
```

```
resuCA2$row$contrib
```

```
##          Dim 1          Dim 2
## 1 66.2331892  4.94824812
## 2  3.0924162 22.39149656
## 3  0.2088724  1.01685164
## 4 16.2001307 10.43299286
## 5  1.9524533  0.01412483
## 6  0.7791956 35.37216674
## 7 10.9724711 23.97662994
## 8  0.5612716  1.84748930
```

```
resuCA2$col$contrib
```

##		Dim 1	Dim 2
##	A	16.138570	3.064199
##	C	59.563097	2.137864
##	D	6.150823	21.603296
##	I	2.997656	69.783959
##	N	15.149854	3.410682

Nous avons 8 profils-lignes, les fortes contributions sont celles > 12.5% (1/8)

Axe 1: 1 4

Axe 2: 2 6 7

Nous avons 5 profils colonnes, les fortes contributions sont celles > 20% (1/5)

Axe1: C

Axe2: D I

4-Profils bien représentés

#Tableau des normes au carré

```
resuCA2$row$cos2
```

##		Dim 1	Dim 2
##	1	0.96653143	0.0273063884
##	2	0.24644364	0.6748003181
##	3	0.10583332	0.1948369997
##	4	0.72484605	0.1765259865
##	5	0.32568515	0.0008909906
##	6	0.05410639	0.9288298783
##	7	0.54643011	0.4515343503
##	8	0.12218869	0.1520940794

```
resuCA2$col$cos2
```

##		Dim 1	Dim 2
##	A	0.78443940	0.05632274
##	C	0.95829798	0.01300696
##	D	0.40023290	0.53158420
##	I	0.09975764	0.87819801
##	N	0.67869354	0.05778019

Profils bien représentés (seuil>0.5 sur l'axe 1 et seuil>0.25 sur l'axe 2)

Axe 1: 1 4 7 A C N

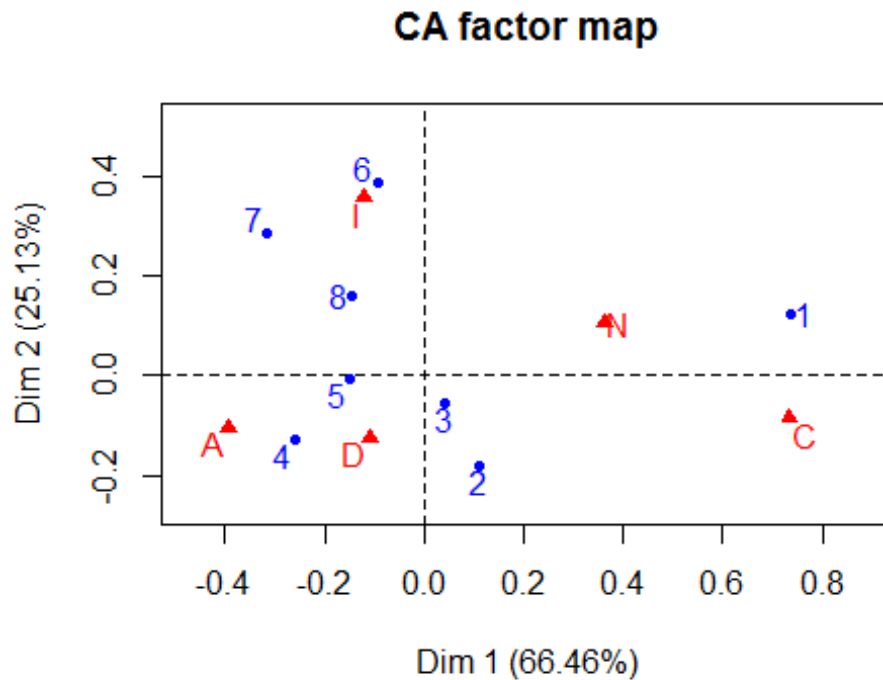
Axe 2: 2 6 7 D I

5-Commentaire sur l'AFC et Graphique

Intersection entre grosse contribution et bonne représentation

Axe1: 1 4 C

Axe2: 2 6 7 D I



Sur l'axe 1, contrairement aux personnes avec un bac+2, les personnes sans diplôme sont le plus touchées par le chômage.

Sur l'Axe 2, on peut remarquer deux groups. D'un coté les personnes avec bac+4 et bac+5 qui sont souvent en CDI et de l'autre les personnes avec un CAP ou BEP qui sont souvent en CDD.

Partie II: AFCM

1-conversion des variables en caractère

```
prison[, ]= lapply(prison[, ], as.character)
```

2-AFCM et commentaires

##	eigenvalue	percentage of variance	cumulative	percentage of variance
## dim 1	0.22264545	22.264545		22.26455
## dim 2	0.13968713	13.968713		36.23326
## dim 3	0.12600462	12.600462		48.83372
## dim 4	0.12169922	12.169922		61.00364
## dim 5	0.10490199	10.490199		71.49384
## dim 6	0.09525251	9.525251		81.01909
## dim 7	0.08654939	8.654939		89.67403
## dim 8	0.08413015	8.413015		98.08705
## dim 9	0.01912953	1.912953		100.00000

Je décide de retenir 5 axes malgré le 5ème axe ayant une valeur propre <0.11. Si on ne retenait pas ce dernier on n'aurait eu que 61% d'inertie totale expliquée alors qu'avec 5 axes on arrive à 71%.

```
resu$var$contrib
```

##	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
## ethnief_0	1.46073963	0.83121002	23.3275707	0.08626739	1.1507810
## ethnief_1	1.55034277	0.88219723	24.7585058	0.09155911	1.2213710
## alcool_0	1.08732119	2.66660587	0.3445301	1.60101181	8.2140549
## alcool_1	4.09808844	10.05037593	1.2985258	6.03417651	30.9585832
## drogue_0	0.32557562	0.01532875	8.1275726	6.07151253	0.8070418
## drogue_1	1.02243805	0.04813843	25.5238383	19.06698490	2.5344349
## suivi_0	5.60079358	8.06331122	0.4331669	0.82296422	2.6424246
## suivi_1	2.46814632	3.55332359	0.1908871	0.36266220	1.1644583
## marie_0	0.03822246	9.30761818	0.1256261	2.17645013	2.2494982
## marie_1	0.11145628	27.14091373	0.3663242	6.34650498	6.5595124
## coupable_0	12.31395018	0.12213042	0.7660034	0.29266503	1.6853636
## coupable_1	26.87912913	0.26658865	1.6720471	0.63883491	3.6788444
## travail_0	4.48360113	6.27372097	0.4670790	0.05164393	12.7309790
## travail_1	5.15747571	7.21664630	0.5372799	0.05940589	14.6444148
## vol_0	7.54291981	2.40600349	1.9140803	2.06383933	0.3684709
## vol_1	22.07533869	7.04148305	5.6018058	6.04009501	1.0783783
## crime_0	0.20166332	0.75211704	0.2421987	2.56809239	0.4428906
## crime_1	3.58279770	13.36228709	4.3029581	45.62532973	7.8684980

Les fortes contributions sont les variables dépassant largement la contribution uniforme $1/18 = 0.056$ Soit 5.6%

Axe 1: coupable_0 coupable_1 vol_0 vol_1

Axe 2: alcool_1 suivi_0 marie_0 marie_1 vol_1 crime_1

Axe 3: ethnie_0 ethnie_1 drogue_0 drogue_1

Axe 4: drogue_0 drogue_1 crime_1

Axe 5: alcool_0 alcool_1 marie_1 travail_0 travail_1 crime_1

##	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
## ethnie_0	0.060336342	0.0215406852	0.545316108	0.001947721	0.02239591
## ethnie_1	0.060336342	0.0215406852	0.545316108	0.001947721	0.02239591
## alcool_0	0.103905708	0.1598758855	0.018632937	0.083627683	0.36983589
## alcool_1	0.103905708	0.1598758855	0.018632937	0.083627683	0.36983589
## drogue_0	0.027011620	0.0007978995	0.381620996	0.275340203	0.03154748
## drogue_1	0.027011620	0.0007978995	0.381620996	0.275340203	0.03154748
## suivi_0	0.161686149	0.1460424970	0.007077033	0.012986083	0.03594146
## suivi_1	0.161686149	0.1460424970	0.007077033	0.012986083	0.03594146
## marie_0	0.002999276	0.4582251830	0.005578921	0.093351331	0.08316745
## marie_1	0.002999276	0.4582251830	0.005578921	0.093351331	0.08316745
## coupable_0	0.785354476	0.0048869148	0.027648506	0.010202654	0.05064445
## coupable_1	0.785354476	0.0048869148	0.027648506	0.010202654	0.05064445
## travail_0	0.193188772	0.1695987653	0.011389848	0.001216321	0.25845599
## travail_1	0.193188772	0.1695987653	0.011389848	0.001216321	0.25845599
## vol_0	0.593493348	0.1187723078	0.085233275	0.088761826	0.01365996
## vol_1	0.593493348	0.1187723078	0.085233275	0.088761826	0.01365996
## crime_0	0.075833373	0.1774440581	0.051543968	0.527859179	0.07846931
## crime_1	0.075833373	0.1774440581	0.051543968	0.527859179	0.07846931

On cherche les variables bien représentées:

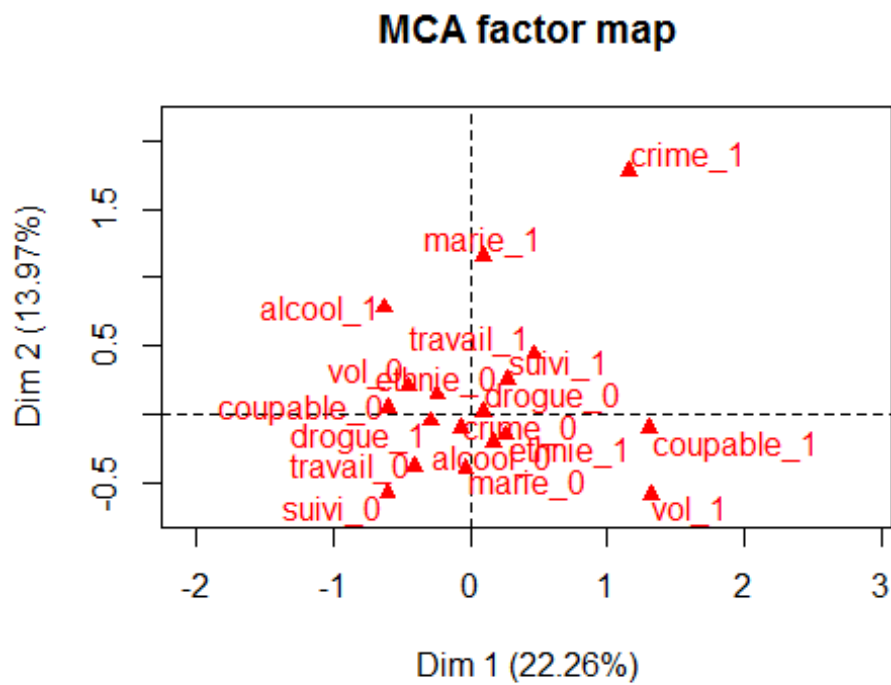
Axe 1 (seuil>30%): coupable_0 coupable_1 vol_0 vol_1

Axe 2 (seuil>30%): marie_0 marie_1

Axe 3 (seuil>20%): ethnie_0 ethnie_1 drogue_0 drogue_1

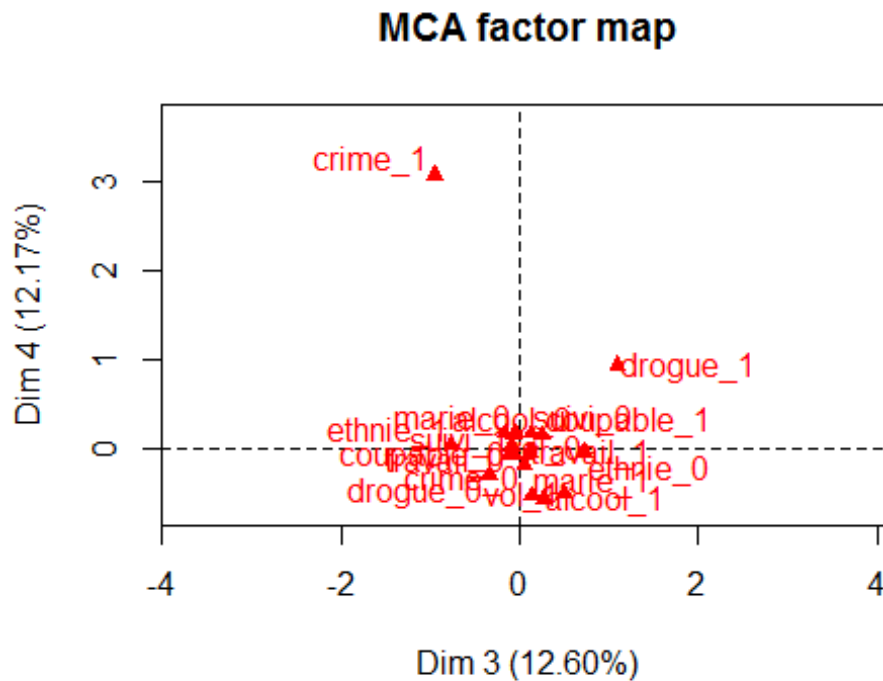
Axe 4 (seuil>10%): drogue_0 drogue_1 crime_0 crime_1

Axe 5 (seuil>10%): alcool_0 alcool_1 travail_0 travail_1



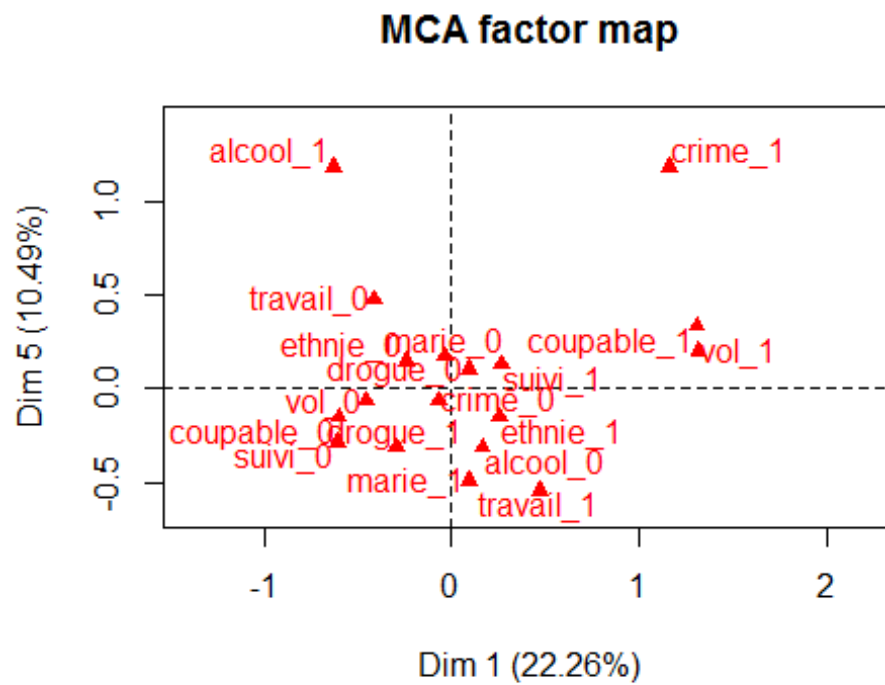
Axe 1: Sur ce premier axe s oppose les individus non coupable et les individus coupable pour vol

Axe 2: Le deuxieme axe oppose les individus mariés aux individus non mariés



Axe 3: Le troisième axe oppose les consommateurs de drogues non caucasien aux non consommateur de drogue caucasien

Axe 4: Cet axe oppose les condamnés pour crime qui consomment de la drogue aux prisonniers qui n'ont pas été condamné pour crime et qui ne consomme pas de la drogue



Axe 5: Ce dernier axe oppose les prisonniers alcooliques ou les prisonniers qui ne travaillent pas avec les prisonniers non alcooliques qui travaillent.

ANNEXE: CODE R

```
library(FactoMineR)

chom <- read.delim("C:/Users/Chris/Desktop/Prepa_M2/Data Mining/Cours 3 - AFC
AFCM/chom.txt")
```

#PARTIE 1: AFC

```
#Création de la table de contingence

chomtable=table(chom$anneetud,chom$emploi)
```

```
#Test d'indépendance

test=chisq.test(chomtable)

test
```

```
test$residuals^2

test$expected

test$observed
```

```
resuCA=CA(chomtable)

resuCA$eig
```

```
resuCA2=CA(chomtable,ncp=2)
```

```
resuCA2$row$contrib

resuCA2$col$contrib

resuCA2$row$cos2

resuCA2$col$cos2
```

#PARTIE 2: AFCM

```
prison = read.csv("C:/Users/Chris/Desktop/Prepa_M2/Data Mining/Cours 3 - AFC  
AFCM/prison.txt", sep="")
```

```
#conversion de chaque variable en caractère
```

```
prison[, ] = lapply(prison[, ], as.character)
```

```
resu=MCA(prison)
```

```
resu$eig
```

```
resu$var$contrib
```

```
resu$var$cos2
```

```
#interpretation de l'AFCM
```

```
plot(resu,axes=(c(1,2)),invisible = "ind")
```

```
plot(resu,axes=(c(3,4)),invisible = "ind")
```

```
plot(resu,axes=(c(1,5)),invisible="ind")
```