TD RD6 - ACM

Exercice 1 Un fabriquant de smartphone a effectué une étude de marché parmi 12 étudiant(e)s de master. La couleur préférée, le système d'exploitation préféré et la filière étaient renseignés. Le tableau codé est :

	Couleur	Système	Faculté
1	Noir	Android	Sciences
2	Noir	OS	DEG
3	Argenté	Android	Sciences
4	Argenté	OS	DEG
5	Argenté	Android	Sciences
6	Argenté	Android	Sciences
7	Argenté	Android	Sciences
8	Bleu	OS	DEG
9	Bleu	Android	Sciences
10	Vert	OS	Sciences
11	Vert	OS	DEG
12	Vert	OS	DEG

	Couleur				Sys		Fac		
	N	A	B	V	An	OS	Sc	DEG	$\sum$
1	1	0	0	0	1	0	1	0	
2	1	0	0	0	0	1	0	1	
3	0	1	0	0	1	0	1	0	
4	0	1	0	0	0	1	0	1	
5	0	1	0	0	1	0	1	0	
6	0	1	0	0	1	0	1	0	
7	0	1	0	0	1	0	1	0	
8	0	0	1	0	0	1	0	1	
9	0	0	1	0	1	0	1	0	
10	0	0	0	1	0	1	1	0	
11	0	0	0	1	0	1	0	1	
12	0	0	0	1	0	1	0	1	
$\sum$	2	5	2	3	6	6	7	5	36

	N	A	B	V	An	OS	Sc	DEG	
	1	2	3	4	5	6	7	8	$\sum$
1	2	0	0	0	1	1	1	1	
2	0	5	0	0	4	1	1	4	
3	0	0	2	0	1	1	1	1	
4	0	0	0	3	0	3	2	1	
5	1	4	1	0	6	0	6	0	
6	1	1	1	3	0	6	5	1	
7	1	4	1	1	6	5	7	0	
8	1	1	1	2	0	1	0	5	
$\sum$									

1. Construire ci-dessus le tableau disjonctif complet puis le tableau de Burt.
2. Calculer les marges des deux tableaux.
3. Calculer l'inertie totale, les inerties des variables et des modalités.

4. Calculer η^2 pour chaque variable.

Exercice 2 Une AFC ou une ACM ?

On considère deux variables qualitatives, X_L et X^C avec un tableau disjonctif complet $Z = (Z_1|Z_2)$ de dimension $n \times p$ avec $p = p_L + p_C = I + J$.

1. A quoi est égal le tableau de contingence entre ces 2 variables, noté N ? On note F la matrice des fréquences relatives, D_I et D_J les matrices diagonales des fréquences marginales.

2. ACM = AFC du tableau Z.
   On pose $D = \begin{pmatrix} D_I & 0 \\ 0 & D_J \end{pmatrix}$ la matrice des fréquences marginales des modalités (D=diag$z_{.j}/n$))

   (a) Construire les profils lignes et colonnes.

   (b) Montrer que la matrice d'inertie des profils colonnes est $\frac{1}{2}\begin{pmatrix} I_I & D_I^{-1}F \\ D_J^{-1}F^T & I_J \end{pmatrix}$.

   (c) On note $\begin{pmatrix} G_I^s \\ G_J^s \end{pmatrix}$ les composantes principales des $I+J$ modalités suivant l'axe s. On pose μ_s la valeur propre de l'axe s. Montrer que l'on a

$$D_I^{-1}FG_J^s = (2\mu_s - 1)G_I^s \text{ et } D_J^{-1}F^TD_I^{-1}FG_J^s = (2\mu_s - 1)^2G_J^s$$
$$D_J^{-1}F^TG_I^s = (2\mu_s - 1)G_J^s \text{ et } D_I^{-1}FD_J^{-1}F^TG_I^s = (2\mu_s - 1)^2G_I^s$$

3. AFC du tableau N.

   On rappelle que dans la DVS de $(X = D_I^{-1}F, Q = D_J^{-1}, D = D_I)$ des profils lignes, nous avons trouvé :

   - la matrice d'inertie $X^T DXQ = F^T D_I^{-1} F D_J^{-1}$,
   - les vecteurs propres V_L vérifiant $X^T DXQV_L = F^T D_I^{-1} F D_J^{-1} V_L = V_L \Sigma^2$, avec $V_L^T D_J^{-1} V_L = I_r$
   - et $F_L = XQV_L = D_I^{-1} F D_J^{-1} V_L$.

   (a) En déduire que $\lambda_s = (2\mu_s - 1)^2$

   (b) En déduire des relations entre F_L et G_I et entre F^C et G_J.

Exercice 3 AFC ou ACM ?

On a observé sur un échantillon de 296 prix nobel la nationalité et la nature du prix. Le tableau de contingence obtenu est :

pays \ nobel	CHIM	LITT	MEDE	PHYS	SECO
FRA	6	11	7	9	0
GB	21	6	19	20	2
RFA	24	7	11	14	0
USA	24	8	55	43	9

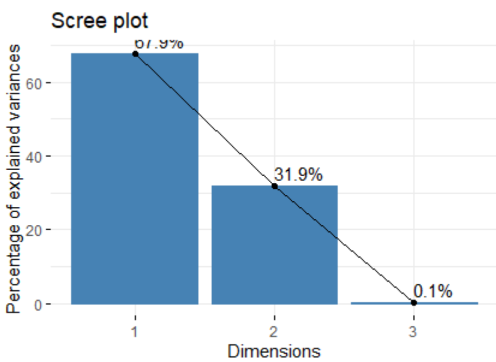
$\left.\begin{array}{l} p=4 \\ n=5 \end{array}\right) \min = 4$

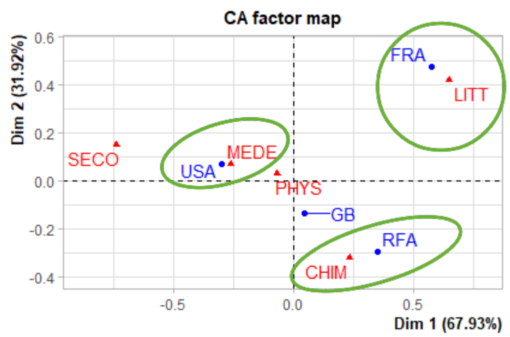
donc $nb\_\lambda = 4-1 = 3 \Rightarrow \underline{3 \text{ dim}}$

On en choisit 2.

1. Une AFC

   Les résultats de l'AFC du tableau de contingence est décrit par les tableaux et diagramme suivants. Interprétez les.





Valeurs propres 0.1024346074 0.0481359059 0.0002180589

	F1	F2	cos2 F1	cos2 F2	contrib F1	contrib F2
CHIM	0.23	−0.32	0.34	0.66	13.41	54.61
LITT	0.65	0.42	0.70	0.30	44.48	39.84
MEDE	−0.26	0.07	0.93	0.07	20.92	3.29
PHYS	−0.07	0.03	0.78	0.13	1.36	0.49
SECO	−0.74	0.15	0.96	0.04	19.82	1.77

	F1	F2	cos2 F1	cos2 F2	contrib F1	contrib F2
FRA	0.58	0.48	0.59	0.41	36.04	52.78
GB	0.04	−0.13	0.09	0.88	0.41	8.65
RFA	0.35	−0.29	0.59	0.41	22.65	33.72
USA	−0.30	0.07	0.95	0.05	40.90	4.85

<u>Axe 1</u> : + Litt | France
− Medecin | USA

<u>Axe 2</u> : + Litt | Fra
− RFA | Chim .

2. Une ACM ?

(a) Le premier prix nobel est américain et en médecine. Quel est la première ligne du tableau codé et du tableau disjonctif complet ?

Codé :

Pays	Nobel
USA	Mede

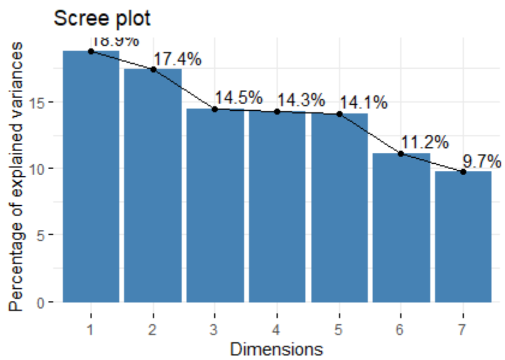
→

USA	FRA	RFA	GB	Mede	chim	Litt	Phys	seco
1	0	0	0	1	0	0	0	0

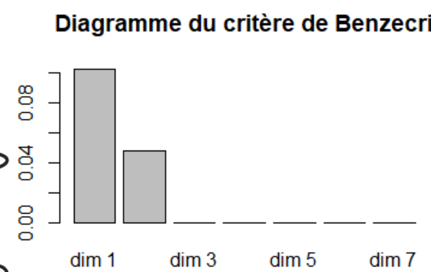
(b) Les résultats de l'ACM sont représentés ci dessous

i. Calculer les valeurs propres avec la correction de Benzécri. Comparez les résultats concernant l'inertie avec ou sans la correction de Benzécri. Comparez à l'AFC.

```
Valeurs propres 0.660 0.610 0.507 0.500 0.493 0.390 0.340
```

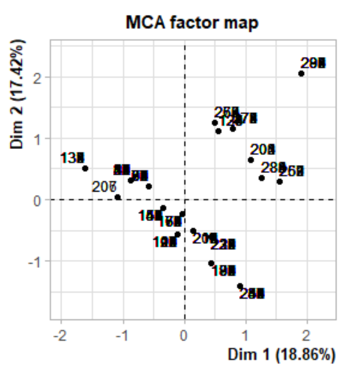


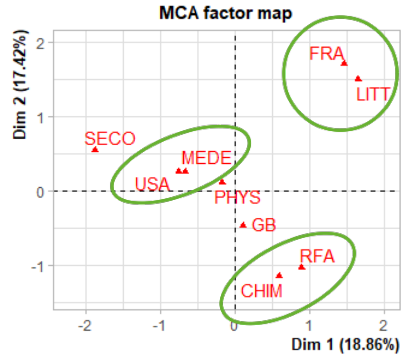
⇒



à faire la transformation soit-même

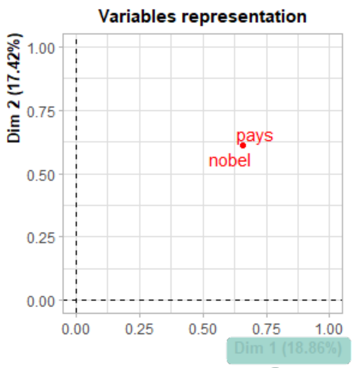
ii. Décrire les résultats concernant les modalités et les variables. Comparer à l'AFC.

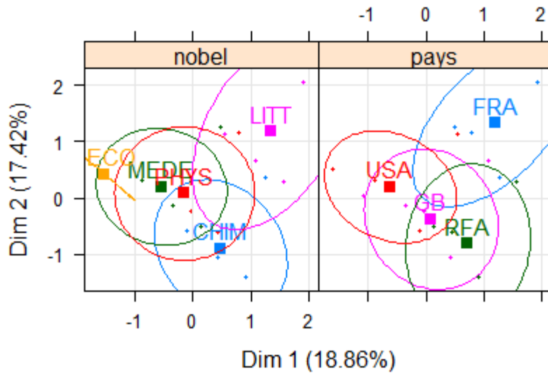




exactement la même que celui de l'AFC

	FRA	GB	RFA	USA	CHIM	LITT	MEDE	PHYS	SECO
contrib F1	18.02	0.21	11.32	20.45	6.71	22.24	10.46	0.68	9.91
contrib F2	26.39	4.33	16.86	2.43	27.30	19.92	1.65	0.24	0.89
v.test F1	8.89	1.02	7.37	−12.25	5.91	9.85	−7.69	−1.93	−6.33
v.test F2	10.34	−4.50	−8.65	4.06	−11.47	8.96	2.93	1.11	1.82





corrélation avec l'axe 1.

η^2

	Dim 1	Dim 2
pays	0.660027	0.6096995
nobel	0.660027	0.6096995

Exercice 4 Gentil toutou ?

Pour le savoir, répondre aux questions suivantes puis en faire une synthèse.

AD9 : caractéristiques de différentes races de chiens.

On utilise des données les caractéristiques de 27 races de chiens au moyen de sept variables : taille, poids, vélocité, intelligence, affection, agressivité et fonction. Les quatre premières variables ont trois modalités chacune (petite, 1 ; moyenne, 2 ; grande, 3), les deux suivantes, deux modalités (faible, 1 ; forte, 2), et la dernière, trois modalités (compagnie, 1 ; chasse, 2 ; utilité, 3). Cette dernière variable sera considérée comme une variable supplémentaire. Les données sont : 6 variables mais 16 modalités -6 $=10$ valeurs propres

	TAI	POI	VEL	INT	AFF	AGR	FON
beaucero	3	2	3	2	2	2	3
basset	1	1	1	1	1	2	2
bergeral	3	2	3	3	2	2	3
boxer	2	2	2	2	2	2	1
buldog	1	1	1	2	2	1	1
bullmast	3	3	1	3	1	2	3
caniche	1	1	2	3	2	1	1
chihuahu	1	1	1	1	2	1	1
cocker	2	1	1	2	2	2	1
colley	3	2	3	2	2	1	1
dalmatie	2	2	2	2	2	1	1
doberman	3	2	3	3	1	2	3
dogueall	3	3	3	1	1	2	3
epagneub	2	2	2	3	2	1	2

	TAI	POI	VEL	INT	AFF	AGR	FON
epagneuf	3	2	2	2	1	1	2
foxhound	3	2	3	1	1	2	2
foxterri	1	1	2	2	2	2	1
gbdeasco	3	2	2	1	1	2	2
labrador	2	2	2	2	2	1	2
levrier	3	2	3	1	1	1	2
mastiff	3	3	1	1	1	2	3
pekinois	1	1	1	1	2	1	1
pointer	3	2	3	3	1	1	2
saintber	3	3	1	2	1	2	3
setter	3	2	3	2	1	1	2
teckel	1	1	1	2	2	1	1
terreneu	3	3	1	2	1	1	3

On fait une analyse des correspondances multiples sur les 6 premières variables. Les valeurs propres associées aux axes sont

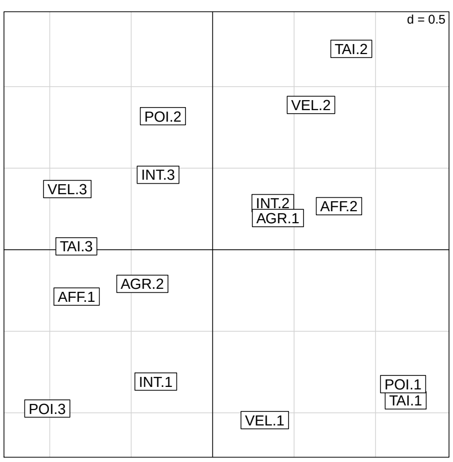
```
[1] 0.48 0.38 0.21 0.16 0.15 0.12 0.08 0.05 0.02 0.01
```

$> 1/6$ $\qquad \frac{1}{6} \approx 0,17$ $\qquad \boxed{\frac{1}{q} \text{ avec } q \text{ nb de question}}$

Les poids n_j/np de chaque catégorie sont (en %) On conserve plutôt 2 axes pour simplifier

TAI.1	TAI.2	TAI.3	POI.1	POI.2	POI.3	VEL.1	VEL.2	VEL.3	INT.1	INT.2	INT.3	AFF.1	AFF.2	AGR.1	AGR.2
4.3	3.1	9.3	4.9	8.6	3.1	6.2	4.9	5.6	4.9	8.0	3.7	8.0	8.6	8.6	8.0

La représentation des catégories sur le premier plan principal, les contributions des catégories aux axes et la qualité de leur représentation (en %) par les sous espaces principaux sont donnés par les deux tableaux ci-dessous.



contrib

	Axis1	Axis2	Axis3
TAI.1	12.6	9.6	7.8
TAI.2	4.6	12.2	15.1
TAI.3	13.5	0.0	0.1
POI.1	14.0	8.7	3.0
POI.2	1.7	15.1	2.2
POI.3	6.6	7.6	21.8
VEL.1	1.3	17.5	4.7
VEL.2	3.7	10.1	3.0
VEL.3	9.2	2.0	15.3
INT.1	1.2	8.4	2.9
INT.2	2.3	1.7	9.3
INT.3	0.9	2.0	6.3
AFF.1	11.6	1.7	0.2
AFF.2	10.8	1.6	0.2
AGR.1	2.9	0.8	3.9
AGR.2	3.1	0.9	4.2

qualité = cos2

	Axis1	Axis1:2	Axis1:3 → axis 1+2
TAI.1	49.1	79.0	92.3
TAI.2	16.5	50.9	74.4
TAI.3	87.5	87.6	87.9
POI.1	57.5	86.1	91.6
POI.2	10.0	82.3	88.0
POI.3	23.4	45.0	78.9
VEL.1	6.0	70.2	79.7
VEL.2	15.3	48.5	53.9
VEL.3	39.8	46.7	75.8
INT.1	5.1	32.7	37.9
INT.2	12.7	20.2	42.8
INT.3	3.2	9.2	19.5
AFF.1	64.8	72.4	72.8
AFF.2	64.8	72.4	72.8
AGR.1	17.3	21.4	31.7
AGR.2	17.3	21.4	31.7

On donne enfin la représentation jointe des chiens et de leurs catégories sur le premier plan principal, ainsi que les coordonnées desdits chiens

coordonnées

	Axis1	Axis2	Axis3
beaucero	-0.32	0.42	-0.10
basset	0.25	-1.10	-0.19
bergeral	-0.49	0.46	-0.50
boxer	0.45	0.88	0.69
buldog	1.01	-0.55	-0.16
bullmast	-0.75	-0.55	0.50
caniche	0.91	0.02	-0.58
chihuahu	0.84	-0.84	-0.47
cocker	0.73	-0.08	0.66
colley	-0.12	0.53	-0.33
dalmatie	0.65	0.99	0.46
doberman	-0.87	0.32	-0.45
dogueall	-1.05	-0.51	0.17
epagneub	0.48	1.04	0.06
epagneuf	-0.14	0.52	0.12
foxhound	-0.88	-0.03	-0.36
foxterri	0.88	-0.14	0.05
gbdeasco	-0.52	0.11	0.04
labrador	0.65	0.99	0.46
levrier	-0.68	0.08	-0.60
mastiff	-0.76	-0.89	0.59
pekinois	0.84	-0.84	-0.47
pointer	-0.67	0.42	-0.69
saintber	-0.58	-0.59	0.89
setter	-0.50	0.38	-0.29
teckel	1.01	-0.55	-0.16
terreneu	-0.38	-0.49	0.66

Question 1 *Y a-t-il des individus identiques ? Est-ce un problème ?*

oui non

Question 2 *Combien de valeurs propres doit on retenir ? Quelle proportion de l'inertie est alors expliquée ?*

3 qui sont > 1/6 mais on en retient 2

$$\frac{\lambda_1+\lambda_2}{\sum \lambda_i} = \frac{0,86}{1,66}$$

Question 3 *Quelles sont les catégories qui déterminent le plus le premier plan principal ?*

Question 4 *Citez 5 catégories qui sont particulièrement mal représentées par le premier plan principal.*

Question 5 *Quels sont les individus qui contribuent le plus au premier plan principal ?*

Question 6 *On donne ci-dessous les coordonnées de la variable supplémentaire FON ainsi que ses valeurs-test.*

	Axis1	Axis2	Axis3
FON.1	1.04	-0.10	-0.07
FON.2	-0.32	0.43	-0.35
FON.3	-0.94	-0.37	0.48

	Axis1	Axis2	Axis3
FON.1	4.06	-0.37	-0.27
FON.2	-1.16	1.56	-1.26
FON.3	-3.10	-1.22	1.58

Expliquez comment les coordonnées de FON ont été calculées. Comment FON se place-t-elle sur les axes ? Expliquez pourquoi les valeurs test ne devraient pas être utilisées. Que disent les valeurs-test tout de même ?