TD 1
Analyse de données

Chazal Baptiste

Exercice 3 :

$$M = \begin{pmatrix} 4 & 3 & 0 & 1 \\ 9 & 7 & 2 & 3 \\ 4 & 0 & 2 & 1 \\ 3 & -1 & 4 & 0 \end{pmatrix}$$

$$\det(M) = \begin{vmatrix} 4 & 3 & 0 & 1 \\ 9 & 7 & 2 & 3 \\ 4 & 0 & 2 & 1 \\ 3 & -1 & 4 & 0 \end{vmatrix} = \begin{vmatrix} 0 & 3 & -2 & 0 \\ -3 & 7 & -4 & 0 \\ 4 & 0 & 2 & 1 \\ 3 & -1 & 4 & 0 \end{vmatrix} = -1 \begin{vmatrix} 0 & 3 & -2 \\ -3 & 7 & -4 \\ 3 & -1 & 4 \end{vmatrix} = -1 \begin{vmatrix} 0 & 3 & -2 \\ 0 & 6 & 0 \\ 3 & -1 & 4 \end{vmatrix} = -1(3) \begin{vmatrix} 3 & -2 \\ 6 & 0 \end{vmatrix}$$

$L_1 \leftarrow L_1 - L_3$
$L_2 \leftarrow L_2 - 3L_3$

$L_2 \leftarrow L_2 + L_3$

$$= -1(3)(0+12)$$
$$= -36$$

Exercice 4 :

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}$$

$$\det(M) = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -2 & -4 \end{vmatrix} = 1 \begin{vmatrix} -1 & -2 \\ -2 & -4 \end{vmatrix} = \left((-1)(-4) - (-2)(-2)\right) = 4 - 4 = 0$$

$L_2 \leftarrow L_2 - 2L_1$
$L_3 \leftarrow L_3 - 3L_1$

Exercice 6 :

$$M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \\ 1 & 5 & 15 & 35 \end{pmatrix}$$

$$\det(M) = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \\ 1 & 5 & 15 & 35 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 3 & 6 \\ 0 & 2 & 7 & 16 \\ 0 & 3 & 12 & 31 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 6 \\ 2 & 7 & 16 \\ 3 & 12 & 31 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 6 \\ 0 & 1 & 4 \\ 0 & 3 & 13 \end{vmatrix} = \begin{vmatrix} 1 & 4 \\ 3 & 13 \end{vmatrix} = 13 - 12 = 1$$

$L_2 \leftarrow L_2 - L_1$
$L_3 \leftarrow L_3 - L_1$
$L_4 \leftarrow L_4 - L_1$

$L_2 \leftarrow L_2 - 2L_1$
$L_3 \leftarrow L_3 - 3L_1$

L2
TD 2
Analyse de données

Exercice 1 :

$$A = \begin{bmatrix} 0 & 1 & 1 \\ -1/2 & 3/2 & -1/2 \\ 3/2 & -3/2 & 1/2 \end{bmatrix} \longrightarrow \begin{vmatrix} -\lambda & 1 & 1 \\ -\frac{1}{2} & \frac{3}{2}-\lambda & -\frac{1}{2} \\ 3/2 & -3/2 & \frac{1}{2}-\lambda \end{vmatrix} = 0$$

$$\xrightarrow{C_2 \leftarrow C_2 - C_3} \begin{vmatrix} -\lambda & 0 & 1 \\ -\frac{1}{2} & 2-\lambda & -\frac{1}{2} \\ \frac{3}{2} & -2+\lambda & \frac{1}{2}-\lambda \end{vmatrix} = 0$$

$$\xrightarrow{L_2 \leftarrow L_2 + L_3} \begin{vmatrix} -\lambda & 0 & 1 \\ 1 & 0 & -\lambda \\ 3/2 & -2+\lambda & \frac{1}{2}-\lambda \end{vmatrix} = 0$$

$$\Rightarrow -(-2+\lambda) \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = 0$$

$$\Rightarrow -(-2+\lambda)(\lambda^2 - 1) = 0$$

$$\Rightarrow \lambda = 2 \text{ ou } \lambda^2 = 1$$

$$\Rightarrow \lambda = 2 \text{ ou } \lambda = 1 \text{ ou } -1 \leftarrow \text{valeurs propres}$$

Pour $\lambda = 1$:

$A\vec{u} = 1\vec{u}$

$$\begin{pmatrix} 0 & 1 & 1 \\ -\frac{1}{2} & \frac{3}{2} & -\frac{1}{2} \\ 3/2 & -\frac{3}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1 \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} y + z = x \\ -\frac{1}{2}x + \frac{3}{2}y - \frac{1}{2}z = y \\ 3/2\,x - \frac{3}{2}y + \frac{1}{2}z = z \end{cases}$$

$$\Rightarrow \begin{cases} -x + y + z = 0 \\ -\frac{1}{2}x + \frac{1}{2}y - \frac{1}{2}z = 0 \\ \frac{3}{2}x - \frac{3}{2}y - \frac{1}{2}z = 0 \end{cases} \Rightarrow \begin{pmatrix} -1 & 1 & 1 & 0 \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & 0 \\ 3/2 & -\frac{3}{2} & -\frac{1}{2} & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} -1 & 1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 3 & -3 & -1 & 0 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} -1 & 1 & 1 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix} \Rightarrow \begin{cases} -x + y + z = 0 \\ z = 0 \end{cases} \Rightarrow \vec{u} = \begin{pmatrix} y \\ y \\ 0 \end{pmatrix}$$

On cherche des vecteurs propres unitaires :

$$\|u\|^2 = y^2 + y^2 + 0^2 = 2y^2 = 1$$

= 1 car on cherche des vecteurs unitaires

$$\Rightarrow y^2 = \frac{1}{2}$$

$$\Rightarrow y = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}} \text{ ou } -\frac{1}{\sqrt{2}}$$

$\rightarrow$ Par exemple $\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$

On fait de même pour $\lambda = -1$ et $\lambda = 2$.

Exercice 3 :

1)

$$X = \begin{pmatrix} 1 & 6 \\ 2 & 8 \\ 3 & 11 \\ 2 & 10 \\ 5 & 7 \\ 6 & 5 \\ 7 & 9 \\ 6 & 8 \end{pmatrix} \qquad \bar{x} = 32/8 = 4 \quad \bar{y} = 64/8 = 8 \qquad X^c = \begin{pmatrix} -3 & -2 \\ -2 & 0 \\ -1 & 3 \\ -2 & 2 \\ 1 & -1 \\ 2 & -3 \\ 3 & 1 \\ 2 & 0 \end{pmatrix}$$

← centrée

$$\Sigma = \frac{1}{8}\begin{bmatrix} -3 & -2 & -1 & -2 & 1 & 2 & 3 & 2 \\ -2 & 0 & 3 & 2 & -1 & -3 & 1 & 0 \end{bmatrix}\begin{bmatrix} -3 & -2 \\ -2 & 0 \\ -1 & 3 \\ -2 & 2 \\ 1 & -1 \\ 2 & -3 \\ 3 & 1 \\ 2 & 0 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 36 & -5 \\ -5 & 28 \end{bmatrix} = \begin{bmatrix} 9/2 & -5/8 \\ -5/8 & 7/2 \end{bmatrix}.$$

2)

$$\det(\Sigma - \lambda I) = \begin{bmatrix} 9/2 - \lambda & -5/8 \\ -5/8 & 7/2 - \lambda \end{bmatrix} = 0$$

$$\Leftrightarrow \left(\frac{9}{2} - \lambda\right)\left(\frac{7}{2} - \lambda\right) - \frac{25}{64} = 0$$

$$\Leftrightarrow \lambda^2 - 8\lambda + \frac{63}{4} - \frac{25}{64} = 0$$

$\Leftrightarrow$ .....

$$\Leftrightarrow \lambda_1 = 4 - \frac{\sqrt{41}}{8} \text{ et } \lambda_2 = 4 + \frac{\sqrt{41}}{8}.$$

3
TD3
Analyse de
données

Exercice 1 :

1)

En cm (la taille) :

$d^2(4,5) = 0^2 + 5^2 + 1^2 + (-1)^2$
$= 0 + 25 + 1 + 1$
$= 27.$

$d^2(4,6) = 0^2 + 5^2 + 1^2 + 2^2$
$= 0 + 25 + 1 + 4$
$= 30.$

$d^2(5,6) = 0^2 + 0^2 + 0^2 + 3^2$
$= 0 + 0 + 0 + 9$
$= 9.$

$A(x, y, z, t)$
$B(x', y', z', t')$
$AB^2 = (x-x')^2 + (y-y')^2 + (z-z')^2 + (t-t')^2$

En m (la taille) :

$d^2(4,5) = 0^2 + 0{,}05^2 + 1^2 + (-1)^2$
$= 0 + 0{,}0025 + 1 + 1$
$= 2{,}0025.$

$d^2(4,6) = 0^2 + 0{,}05^2 + 1^2 + 2^2$
$= 0 + 0{,}0025 + 1 + 4$
$= 5{,}0025.$

$d^2(5,6) = 0^2 + 0^2 + 0^2 + 3^2$
$= 0 + 0 + 0 + 9$
$= 9.$

l'échelle de l'unité de mesure a une grande importance sur l'impact des données concernées sur les distances.
=> effet d'échelle

Exercice 2 :

a)

$AB^2 = 50^2 + (-5)^2 + 0^2 + 0^2 + 3^2 = 2500 + 25 + 9 = 2534.$
$AC^2 = 47^2 + (-10)^2 + 20^2 + (-10)^2 + 3^2 = 2209 + 100 + 400 + 100 + 9 = 2818$
$BC^2 = (-3)^2 + (-5)^2 + 20^2 + (-10)^2 + 0^2 = 9 + 25 + 400 + 100 + 0 = 534.$

b)

Centrer les données ne modifie pas les distances, cela déplace juste le nuage de points. (les soustractions de moyenne s'annulent)

TD 4

ACP du TD 4 :

① Pourquoi une ACP ? Quel type d'ACP ? centrée ? / centrée réduite ?

② Analyse de la matrice des corrélations

③ Choix des axes + qge

④ Cercles de corrélation + analyse des cercles + matrice de saturation

⑤ Choix des individus intéressants + représentation + analyse des individus

⑥ Synthèse

1) Variables quantitatives de même importance

2)

Forte corrélation	Forte anticorrélation
PI - CPIS	PI - Agri
Vieux - Agri	Vieux - Jeunes
CPIS - Pop	PI - Vieux
ACCE - Vieux	CPIS - Vieux
PI - Pop	Pop - Vieux
PI - Jeune	Agri - Pop
	OUV - CPIS

3)

Axe 1 : $\frac{\lambda_1}{\Sigma_y} \times 100 = 47,25\,\%$

Axe 2 : 19,27 %

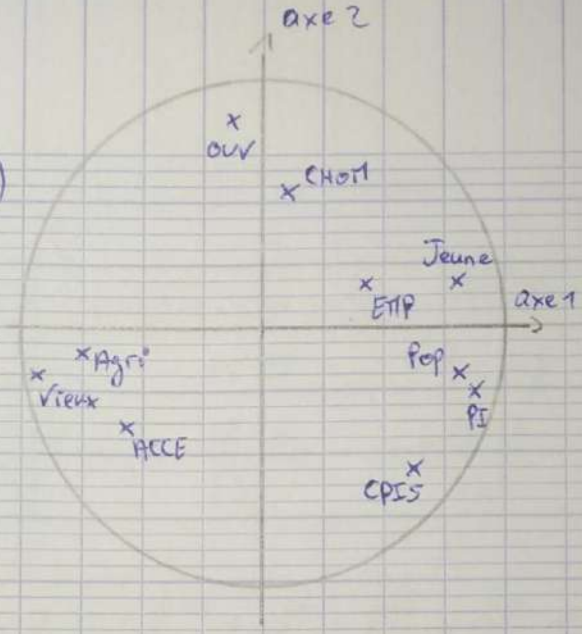
Axe 3 : 12,49 %

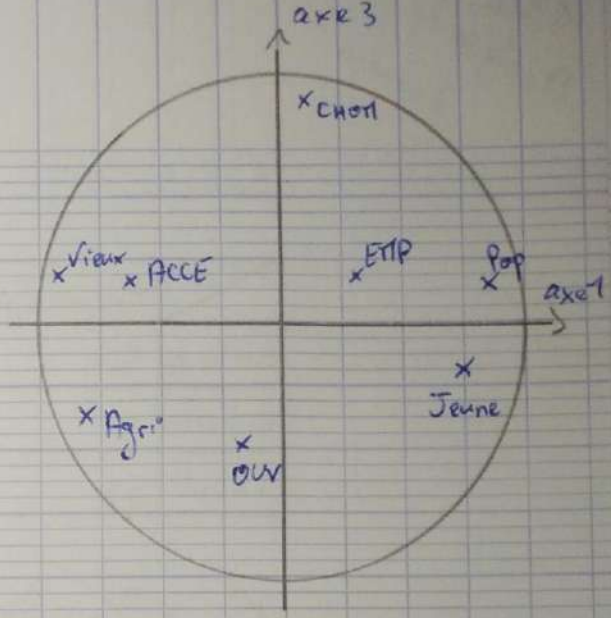
qge = 79 %

4

TD4 Analyse de données

4)





Axe 1 : Pop, PI, CPIS, jeune − Vieux, agris

Axe 2 : OUV

Axe 3 : CHOM

Axe 1

⊖	⊕
12 Aveyron	69 Rhône
15 Cantal	77 Seine & Marne
19 Corrèze	78 Yvelines
23 Creuse	91 Essonne
24 Dordogne	94 Val de Marne
32 Gers	95 Val d'Oise
46 Lot	

Axe 2

64 Pyr Atl	02 Aisne
75	08 Ardennes
06 Alpes Mar	88 Vosges
	55 Meuse
	70 Haute-Saône
	72 Sarthe
	10 Aube
	80 Somme

Axe 1 − Vieux, Agri | Pop, CPIS, PI, Jeune

+ Vieux + Agris	− Vieux − Agri
− Pop − Jeune − CPIS − PI	+ Pop + Jeune + CPIS + PI

→ URBANISATION

Axe 2 Monde ouvrier

Axe 3

53 Mayenne	11 Aude
79 Deux Sèvres	30 Gard
49 Maine-et-Loire	34 Hérault
35 Ille-et-Vilaine	66 Pyrénées-Or
	83 Var
	84 Vaucluse

TD6
Analyse de données

Exercice 1 :

$$\chi^2_{obs} = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(h_{ij} - h'_{ij})^2}{h'_{ij}}$$

	Examinateur A	Examinateur B	Examinateur C	K
Brillants	50 47	47 52	56 54	153
Médiocres	5 3	14 9	8 10	27
Total	55	61	64	180

où $h'_{ij} = \frac{h_{i\cdot} \cdot h_{\cdot j}}{N}$

$$\chi^2_{obs} = 9/47 + 9/8 + 25/52 + 25/9 + 4/54 + \cancel{10}\, 4/10 \simeq 5{,}05$$

H_0 : " Il y a indépendance entre le choix du correcteur et l'évaluation. "

$\nu = (n-1)(p-1) = 2$

Au seuil de 5%.

$\chi^2_{0,95} = 5{,}99$

$\chi^2_{0,95} > \chi^2_{obs}$

$\rightarrow$ On accepte H_0 au seuil de 5%.

Au seuil de 10%.

$\chi^2_{0,90} = 4{,}61$

$\chi^2_{0,90} < \chi^2_{obs}$

$\rightarrow$ On rejette au seuil de 10%

TD 8
Analyse de données

AFC

① Test d'indépendance ($N \simeq 40\,000\,000$)

② Analyse des PFL et PFC

③ Choix des axes + qqe

④ Représentations graphiques avec selection

⑤ Fin de l'analyse

H_0 : il y a indépendance entre la région et le vote

$\chi^2_{obs} = NR$ où $R = \sum \lambda_i - 1 = 0{,}02$
$\chi^2_{obs} \simeq 800\,000$
$\nu = (n-1)(p-1) = 105$
$\chi^2_{obs} >>> \chi^2_{th}$ => on rejette H_0

TD 9 Analyse de données

	Âge		Sexe		Taille		Cheveux		
	Vieux	Jeune	Homme	Femme	Petit	Grand	Blanc	Brun	Blond
Alain	1	0	1	0	1	0	1	0	0
Brigitte	0	1	0	1	0	1	0	1	0
Charles	1	0	1	0	1	0	0	1	0
Diane	1	0	0	1	0	1	1	0	0
Elisabeth	1	0	0	1	1	0	0	1	0
François	0	1	1	0	0	1	0	0	1
Gaston	0	1	1	0	0	1	0	1	0
Henri	1	0	1	0	1	0	0	0	1
Isabelle	0	1	0	1	1	0	0	0	1
Jean	1	0	1	0	0	1	0	1	0
Karl	0	1	1	0	0	1	0	1	0

$R = \frac{m}{q} - 1 = \frac{9}{4} - 1 = 1,25$ m : nb de modalité

Il y a 9 λ en tout, avec $\lambda_6 = 1$ valeur triviale et le reste proche de 0 donc noté nuls.

axe 1 : $\frac{0,416}{1,25} \times 100 = 33\%$

axe 2 : 24 % → qqe de 57% (somme des deux)

→ faire comme une AFC