MAT 3502 2° session

| INT | Ingénieur 1°année | Août 2008 |
|---|---|---|

M. BECKER – B. MORANDO

## Analyse des données

(calculettes et tout document autre que le polycopié de cours sont interdits)

| | X | Y | Z |
|---|---|---|---|
| A | 1 | 0 | 0 |
| B | 0 | 1 | 0 |
| C | 0 | 0 | 1 |
| D | 1 | 1 | 0 |
| E | 1 | 0 | 1 |
| F | 0 | 1 | 1 |

On considère le tableau de données logiques ci-contre :

Selon qu'un individu (A ou B ou C …etc…) est concerné par une variable (X ou Y ou Z) , on note 1 ou 0 dans la case correspondante.

**Première partie : Analyse factorielle des correspondances :**

**1°** - Former la matrice à diagonaliser pour effectuer l'A.F.C. de ce tableau et calculer ses valeurs propres.

**2°** - Calculer les vecteurs propres correspondant aux deux axes factoriels.

**3°** - Calculer les coordonnées dans ces axes des individus et des variables.

**4°** - Dessiner le graphique de cette A.F.C.- Commentaire.

**Deuxième partie : Analyse en composantes principales :**

**1°** - Dresser le tableau des variables centrées réduites ; en déduire la matrice à diagonaliser.

**2°** - Calculer les valeurs propres et les vecteurs propres correspondant aux deux premiers axes.

**3°** - Calculer les coordonnées dans ces axes des individus et des variables.

**4°** - Dessiner le graphique de cette A.C.P.- Commentaire (différence avec l'A.F.C.)

# I: AFC du tableau :

| | X | Y | Z | |
|---|---|---|---|---|
| A | 1 | 0 | 0 | 1 |
| B | 0 | 1 | 0 | 1 |
| C | 0 | 0 | 1 | 1 |
| E | 1 | 1 | 0 | 2 |
| F | 1 | 0 | 1 | 2 |
| G | 0 | 1 | 1 | 2 |
| | 3 | 3 | 3 | |

## 1) Matrice à diagonaliser :

On forme d'abord :

$$S = \left\{ k_{ij} / \sqrt{k_{i.} k_{.j}} \;/\; i, j \right\}$$

$$\Rightarrow S = \begin{pmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{6} & 0 \\ 1/\sqrt{6} & 0 & 1/\sqrt{6} \\ 0 & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix}$$

puis

$${}^tSS = \begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 2/3 & 1/6 \\ 1/6 & 1/6 & 2/3 \end{pmatrix}$$

ou encore :

$${}^tSS = \frac{1}{6} \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}$$

, le calcul des valeurs propres ici les solutions de $\det({}^tSS - \lambda Id) =$ revient à calculer, ayant posé $4 - 6\lambda = X$, les racines d

$$\begin{vmatrix} X & 1 & 1 \\ 1 & X & 1 \\ 1 & 1 & X \end{vmatrix} = 0$$

ie : $X = -2$ et $X = 1$ val. double :

$\Rightarrow \lambda = 1$ val propre triviale

et $\lambda = 1/2$ val propre double.

## 2) Vecteurs propres:

On doit résoudre: ${}^tSS\,u = \lambda u$

et donc: $\frac{1}{6}\begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{3}{6}\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} x+y+z=0 \\ x+y+z=0 \\ x+y+z=0 \end{cases}$

il y a un plan propre ce qui est logique:
Il suffit de choisir deux vecteurs unitaires orthogonaux dt les cosinus directeurs satisfont $x+y+z=0$

par exemple: ${}^1u \left| \begin{matrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{matrix} \right.$ ${}^2u \left| \begin{matrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -\sqrt{2/3} \end{matrix} \right.$

## 3) Coordonnées des i et des j:

On forme la matrice T (de coeff: ${}^\alpha u_j / \sqrt{f_{.j}}$) ici les $\sqrt{f_{.j}} = \frac{1}{\sqrt{2}}$ $\forall j$

$\Rightarrow T = \begin{pmatrix} 1 & 1/\sqrt{3} \\ -1 & 1/\sqrt{3} \\ 0 & -2/\sqrt{3} \end{pmatrix}$ On a donc les coordonnées des individus et des variables en calculant:

$$F = f^I_J \cdot T \quad \text{et} \quad G = T\,\Lambda^{1/2}$$

$\Rightarrow$

$$F = \begin{pmatrix} 1 & 1/\sqrt{3} \\ -1 & 1/\sqrt{3} \\ 0 & -2/\sqrt{3} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1/\sqrt{3} \\ -1 & 1/\sqrt{3} \\ 0 & -2/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 1/2 & -1/2\sqrt{3} \\ -1/2 & -1/2\sqrt{3} \end{pmatrix}$$

$$\begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix}$$

$$\text{et } G = \begin{pmatrix} 1 & 1/\sqrt{3} \\ -1 & 1/\sqrt{3} \\ 0 & -2/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -2/\sqrt{6} \end{pmatrix}$$

Pour le graphique :

$1/\sqrt{3} = 0{,}577$ , $2/\sqrt{3} = 1{,}154\ldots$, $\frac{1}{2\sqrt{3}} = 0{,}288$

$1/\sqrt{2} = 0{,}707$ , $1/\sqrt{6} = 0{,}408\ldots$, $\frac{2}{\sqrt{6}} = 0{,}816$

## 4) Graphique :

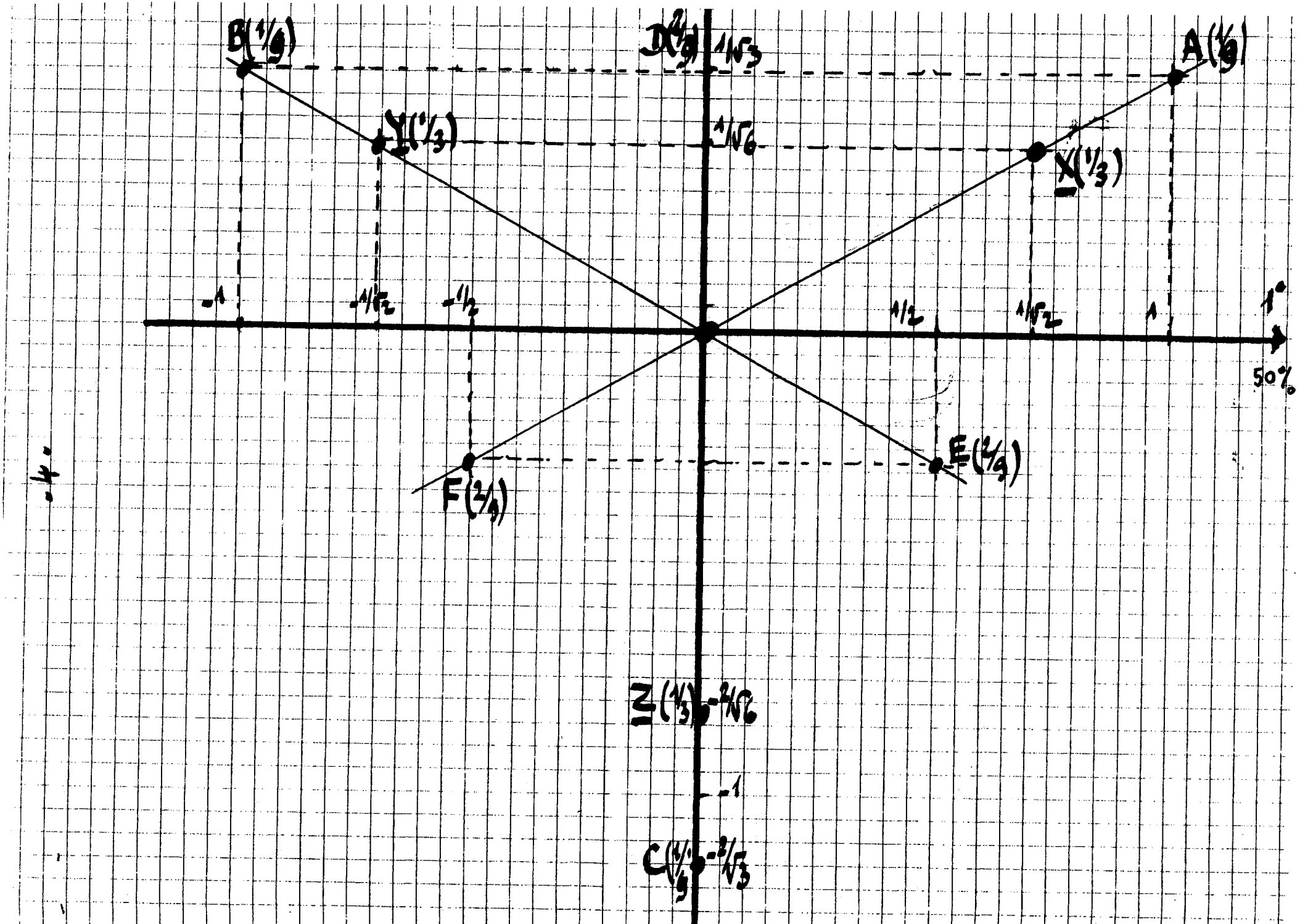
(voir page suivante)

Commentaire : Vérifier que les résultats sont cohérents i.e. la position de individus par rapport aux variables et inversement.

De plus, on a fait figurer les poids relatifs entre parenthèses et on retrouve bien les équilibres par rapport au centre de gravité.

Enfin la symétrie des nuages de A, B, C d'une part, de DEF d'autre part, leurs antisymétries d'autre part se traduisent par :

ABC, DEF, XYZ sont trois triangles équilatéraux. (et DEF s'obtient à CBA) cf tableau.


B(1/3)
D(2/3)
A(1/3)
Y(1/3)
X(1/3)
1/√3
1/√6
-1
-1/√2
-1/2
1/2
1/√2
1
1°
50%
F(2/3)
E(2/3)
Z(1/3)
-2/√6
-1
C(1/3)
-2/√3

## II A.C.P. du tableau précédent :

On forme le tableau $\mathcal{X}$ des variables centrées réduites.
ici : $\bar{X} = \bar{Y} = \bar{Z} = 1/2$ et $\sigma_X = \sigma_Y = \sigma_Z = 1/2$

On a donc immédiatement :

1) Matrice à diagonaliser :

$$\mathcal{X} = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

et donc :

$$^t\mathcal{X}\mathcal{X} = \begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 6 \end{pmatrix}$$

2) Valeurs propres, vecteurs propres :

Posant $6 - \lambda = X$ on résout :

$$\begin{vmatrix} X & -2 & -2 \\ -2 & X & -2 \\ -2 & -2 & X \end{vmatrix} = 0$$

$X = -2$ rend les trois colonnes identiques $X = -2$ est donc racine double
$\Rightarrow \lambda = 8$ val propre double $(\lambda_1 = \lambda_2 = 8)$

La somme des valeurs propres (trace de la matrice) valant 18, la troisième valeur propre vaut 2.

Le taux d'inertie pour le troisième axe étant faible, on se contente de $\lambda_1$ et $\lambda_2$ ; on aura donc aussi ici un plan propre donné par :

$$\begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 6 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 8\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow 3 \text{ fois l'équation: } x + y + z = 0$$

Là encore, on peut prendre par exemple:

$$\begin{array}{c|c} {}^1u & \begin{matrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{matrix} \end{array} \qquad \begin{array}{c|c} {}^2u & \begin{matrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \text{ ou } -\sqrt{2/3} \end{matrix} \end{array}$$

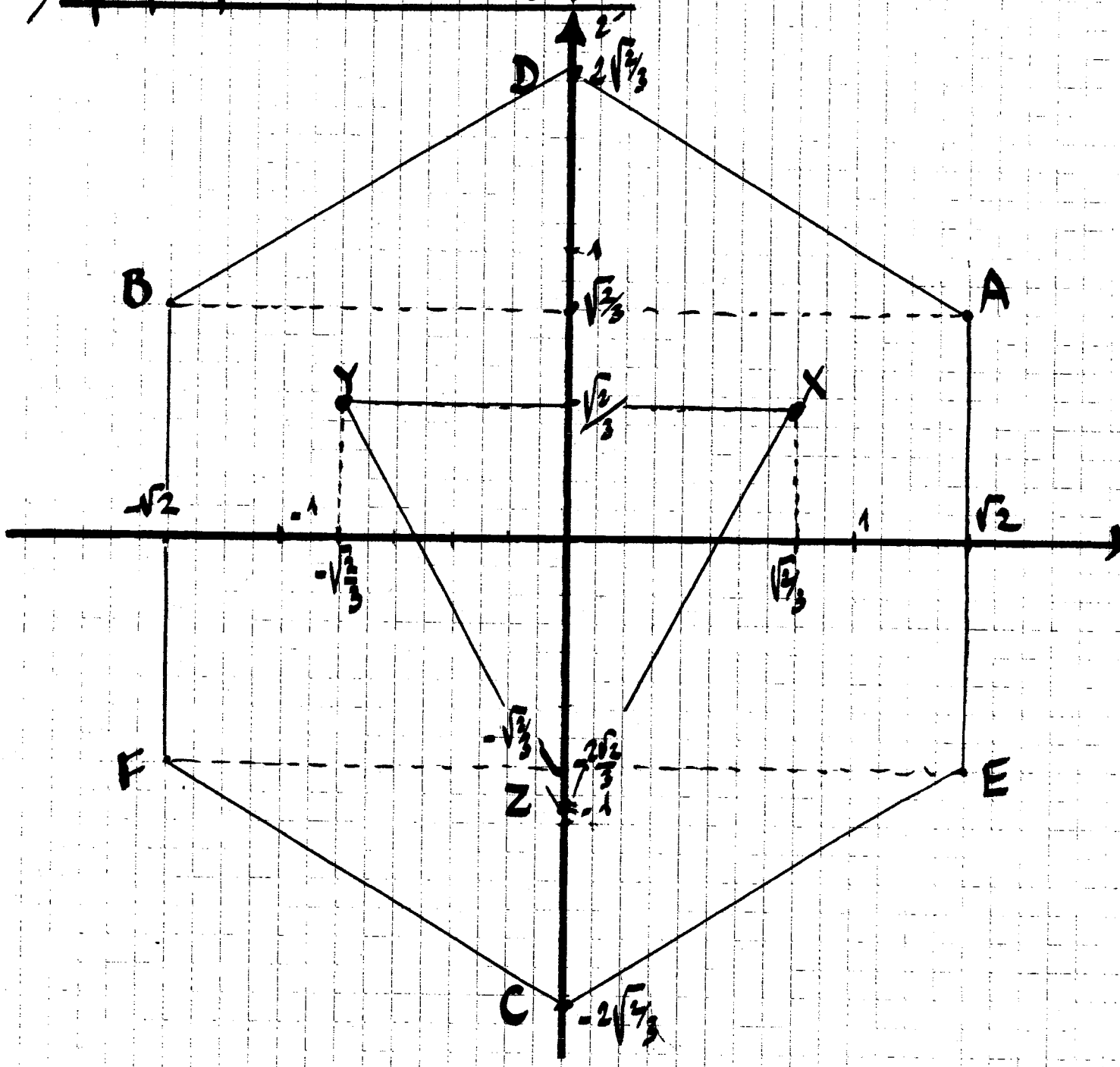
## 3) Coordonnées des points ds le premier plan de l'ACP.

Elles sont données par $F_\alpha(i) = \sum \{ x_{ij}\, {}^\alpha u_j \,/\, j \in J \}$
et $G_\alpha(j) = \sqrt{\frac{\lambda_\alpha}{n}}\, {}^\alpha u_j$ , ce qui donne: pr les deux premiers axes

$$F = \begin{pmatrix} \sqrt{2} & 2/\sqrt{6} \\ -\sqrt{2} & 2/\sqrt{6} \\ 0 & -4/\sqrt{6} \\ 0 & 4/\sqrt{6} \\ \sqrt{2} & -2/\sqrt{6} \\ -\sqrt{2} & -2/\sqrt{6} \end{pmatrix} \qquad G = \begin{pmatrix} \sqrt{\frac{2}{3}} & \frac{\sqrt{2}}{3} \\ -\sqrt{\frac{2}{3}} & \frac{\sqrt{2}}{3} \\ 0 & -\frac{2\sqrt{2}}{3} \end{pmatrix}$$

$\sqrt{2} = 1{,}414$ ; $2/\sqrt{6} = 0{,}816$ ; $4/\sqrt{6} = 1{,}633$ ; $\frac{\sqrt{2}}{3} = 0{,}471$ ; $\frac{2\sqrt{2}}{3} = 0{,}942$
ou $\sqrt{\frac{2}{3}}$ ou $2\sqrt{\frac{2}{3}}$

## 4) Graphique et commentaire :




Mêmes remarques sur la cohérence des résultats que pour l'A.F.C. en remarquant de plus que les poids des lignes étant cette fois identiques, AECFBD forment un hexagone régulier. et toujours X42 un triangle équilatéral

**Contrôle écrit première session d'Analyse des données** (AD12)
Le 14 juin 2007
Monique BECKER et Bertrand MORANDO
(aucun document autre que le poly de cours non annoté n'est autorisé
Les calculettes et téléphones portables sont interdits)

Soit le tableau de données ci-dessous :

| | X | Y | Z | T | U | V |
|---|---|---|---|---|---|---|
| A | 2 | 2 | 2 | 2 | 0 | 0 |
| B | 4 | 1 | 0 | 2 | 1 | 0 |
| C | 0 | 0 | 1 | 2 | 2 | 3 |

**Première partie : Analyse factorielle des correspondances :**

1) En vue de faire l'analyse factorielle des correspondances des données ci-dessus, calculer la matrice à diagonaliser.
2) Calculer les valeurs propres et les vecteurs propres de la sus-dite.
3) Calculer les coordonnées des représentants des lignes et des colonnes.
4) Faire le graphique (lignes et colonnes) du plan factoriel. (en grand format pour la suite) ;
5) Donner les tableaux des contributions relatives des lignes et des colonnes.

**Deuxième partie : Classification ascendante hiérarchique :**

1) On désire faire la classification des colonnes du tableau suite à cette A.F.C. Pour cela, calculer les distances des points deux à deux, puis les indices de niveau associés.
2) En déduire les premiers éléments à agréger, puis continuer l'algorithme .
3) Dessiner le dendrogramme correspondant.
4) Reporter cette classification sur le graphique du plan factoriel de la première partie.
5) Dessiner sans calcul le dendrogramme de la classification des lignes ; remarques .

# INT Ana Don. Juin 2007

## Premiere partie : AFC

| | X | Y | Z | T | U | V | $k_{i\cdot}$ |
|---|---|---|---|---|---|---|---|
| A | 2 | 2 | 2 | 2 | 0 | 0 | 8 |
| B | 4 | 1 | 0 | 2 | 1 | 0 | 8 |
| C | 0 | 0 | 1 | 2 | 2 | 3 | 8 |
| $k_{\cdot j}$ | 6 | 3 | 3 | 6 | 3 | 3 | 24 |

1) On forme la matrice $S = \left\{ s_{ij} = k_{ij} / \sqrt{k_{i\cdot} k_{\cdot j}} \right\}$ puis on diagonalise ${}^tSS$.

Ici, à l'évidence, on a intérêt à raisonner sur $S{}^tS$ et à faire les calculs en commençant par les $j$ de $\mathbb{R}^I$

$$S = \frac{1}{2\sqrt{6}} \begin{pmatrix} \sqrt{2} & 2 & 2 & \sqrt{2} & 0 & 0 \\ 2\sqrt{2} & 1 & 0 & \sqrt{2} & 1 & 0 \\ 0 & 0 & 1 & \sqrt{2} & 2 & 3 \end{pmatrix} \Rightarrow S{}^tS = \frac{1}{24} \begin{pmatrix} 12 & 8 & 4 \\ 8 & 12 & 4 \\ 4 & 4 & 16 \end{pmatrix}$$

ou encore $\boxed{\frac{1}{6} \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}}$

2) on doit de diagonaliser $S{}^tS$ c.à.d trouver les solutions de $\det(S{}^tS - \lambda I) = 0$

ou encore posant $3 - 6\lambda = X$, on résout

$$\begin{vmatrix} X & 2 & 1 \\ 2 & X & 1 \\ 1 & 1 & X+1 \end{vmatrix} = 0$$

$X = 2$ est solution évidente $\Rightarrow \lambda = \frac{1}{6}$

on doit avoir $\lambda_0 = 1$ valeur propre et la somme des vleurs propres valant $\frac{10}{6}$ ; on a $1 + \frac{1}{6} + \lambda = \frac{10}{6}$

Le 3e vleur prope vaut donc $\frac{3}{6} = \frac{1}{2}$.

En résumant : on a : $\lambda_0 = 1$ v.p triviale

$$\boxed{\lambda_1 = \frac{1}{2}, \ \lambda_2 = \frac{1}{6}}$$

vecteurs propres ${}^{\alpha}w$

Pour $\lambda_1 = \frac{1}{2}$ il vient $\frac{1}{6}\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$

$$\Rightarrow \begin{cases} 3x + 2y + z = 3x \\ 2x + 3y + z = 3y \\ x + y + 4z = 3z \end{cases}$$

d'où, en normant à 1 : ${}^{1}w \begin{vmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{vmatrix}$ pour $\lambda_1 = 1/2$

Pour $\lambda_2 = \frac{1}{6}$ il vient $\frac{1}{6}\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{1}{6}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$

$$\Rightarrow \begin{cases} 3x + 2y + z = x \\ 2x + 3y + z = y \\ x + y + 4z = z \end{cases}$$

d'où, en normant à 1 : ${}^{2}w \begin{vmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{vmatrix}$ pour $\lambda_2 = \frac{1}{6}$

3) <u>Coordonnées des colonnes :</u> on forme la matrice $T'$ à partir des ${}^{1}w, {}^{2}w$ en multipliant par : $\frac{1}{\sqrt{f}\,i}$ ici : ici $\times\sqrt{3}$

et on premultiplie $T'$ par : ${}^{t}f^{J}_{I}$

$$\begin{pmatrix} 1/\sqrt{2} & \sqrt{3}/2 \\ 1/\sqrt{2} & -\sqrt{3}/2 \\ -\sqrt{2} & 0 \end{pmatrix}$$

on a donc :

$$\begin{pmatrix} 1/3 & 2/3 & 0 \\ 2/3 & 1/3 & 0 \\ 2/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{6} \\ 0 & 2/\sqrt{6} \\ 0 & 0 \\ -1/\sqrt{2} & -1/\sqrt{6} \\ -\sqrt{2} & 0 \end{pmatrix}$$

Coordonnées des 1-2 de : $\begin{matrix} X \\ Y \\ Z \\ T \\ U \\ V \end{matrix}$

Et l'on a directement les coordonnées de $i$

par: $\begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} T' = \begin{pmatrix} \sqrt{\frac{1}{2}} & 0 \\ 0 & \sqrt{\frac{1}{6}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} \\ \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} \\ -\sqrt{2} & 0 \end{pmatrix}$

$\Rightarrow$

| | | |
|---|---|---|
| Coordonnées de 1.2 de C | A | $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 \end{pmatrix}$ |
| | B | |
| | C | |

4) Graphique: voir feuille jointe:

5) Contributions $CTR_\alpha(i) = \frac{f_{i.} F_\alpha^2(i)}{\lambda_\alpha}$ ; $CTR_\alpha(j) = \frac{f_{.j} G_\alpha^2(j)}{\lambda_\alpha}$

D'où les tableaux de contribution:

| | Axe 1 | Axe 2 |
|---|---|---|
| A | 1/6 | 1/2 |
| B | 1/6 | 1/2 |
| C | 2/3 | 0 |

| | Axe 1 | Axe 2 |
|---|---|---|
| X | 1/4 | 1/4 |
| Y | 1/8 | 1/8 |
| Z | 0 | 1/2 |
| T | 0 | 0 |
| U | 1/8 | 1/8 |
| V | 1/4 | 0 |

## Deuxième partie: C.A.H:

| $d^2$ | X | Y | Z | T | U | V |
|---|---|---|---|---|---|---|
| X | 0 | 2/3 | 2 | 2/3 | 2 | 14/3 |
| Y | | 0 | 2/3 | 2/3 | 8/3 | 14/3 |
| Z | | | 0 | 2/3 | 2 | 8/3 |
| T | | | | 0 | 2/3 | 2 |
| U | | | | | 0 | 2/3 |
| V | | | | | | 0 |

| $\delta$ | X | Y | Z | T | U | V |
|---|---|---|---|---|---|---|
| X | 0 | 1/18 | 1/6 | 1/12 | 1/6 | 7/18 |
| Y | | 0 | 1/24 | 1/18 | 1/6 | 7/24 |
| Z | | | 0 | 1/18 | 1/8 | 1/6 |
| T | | | | 0 | 1/18 | 1/6 |
| U | | | | | 0 | (1/24) |
| V | | | | | | 0 |

On agrège par les U et V (on aurait pu s'embêter à agréger 42 d'abord) dont le barycentre sera le noeud $n_7$

| $d^2$ | X | Y | Z | T | $n_7$ (6) |
|---|---|---|---|---|---|
| X | 0 | 2/3 | 2 | 2/3 | 19/6 |
| Y | | 0 | 2/3 | 2/3 | 21/6 (7/2) |
| Z | | | 0 | 2/3 | 13/6 |
| T | | | | 0 | 7/6 |
| $n_7$ | | | | | 0 |

| $\delta$ | X | Y | Z | T | $n_7$ (6) |
|---|---|---|---|---|---|
| X | 0 | 1/18 | 1/6 | 1/12 | 19/48 |
| Y | | 0 | (1/24) | 1/18 | 7/24 |
| Z | | | 0 | 1/18 | 13/48 |
| T | | | | 0 | 7/48 |
| $n_7$ | | | | | 0 |

On agrège 42 → $n_8$ avec niveau de noeud 1/24 comme le précédent ce qui est logique, puisqu'on aurait pu commencer par là.

| $d^2$ | X | $n_8$ | T | $n_7$ |
|---|---|---|---|---|
| X | 0 | 7/6 | 2/3 | 19/6 |
| $n_8$ | " | 0 | 1/2 | 8/3 |
| T | " | " | 0 | 7/6 |
| $n_7$ | " | " | " | 0 |

| $\delta$ | X | $n_8$ | T | $n_7$ |
|---|---|---|---|---|
| X | 0 | 7/48 | 1/12 | 19/48 |
| $n_8$ | " | 0 | (1/16) | 1/3 |
| T | " | " | 0 | 7/48 |
| $n_7$ | " | " | " | 0 |

On agrège $n_8$ et T $\rightsquigarrow$ on fait le nuage $n_9$ (barycentre de $n_8$, T)

| $d^2$ | X | $n_9$ | $n_7$ |
|---|---|---|---|
| X | 0 | 19/24 | 19/6 |
| $n_9$ | " | 0 | 43/24 |
| $n_7$ | " | " | 0 |

| $\delta$ | X | $n_9$ | $n_7$ |
|---|---|---|---|
| X | 0 | (19/144) | 19/48 |
| $n_9$ | " | 0 | 43/144 |
| $n_7$ | " | " | 0 |

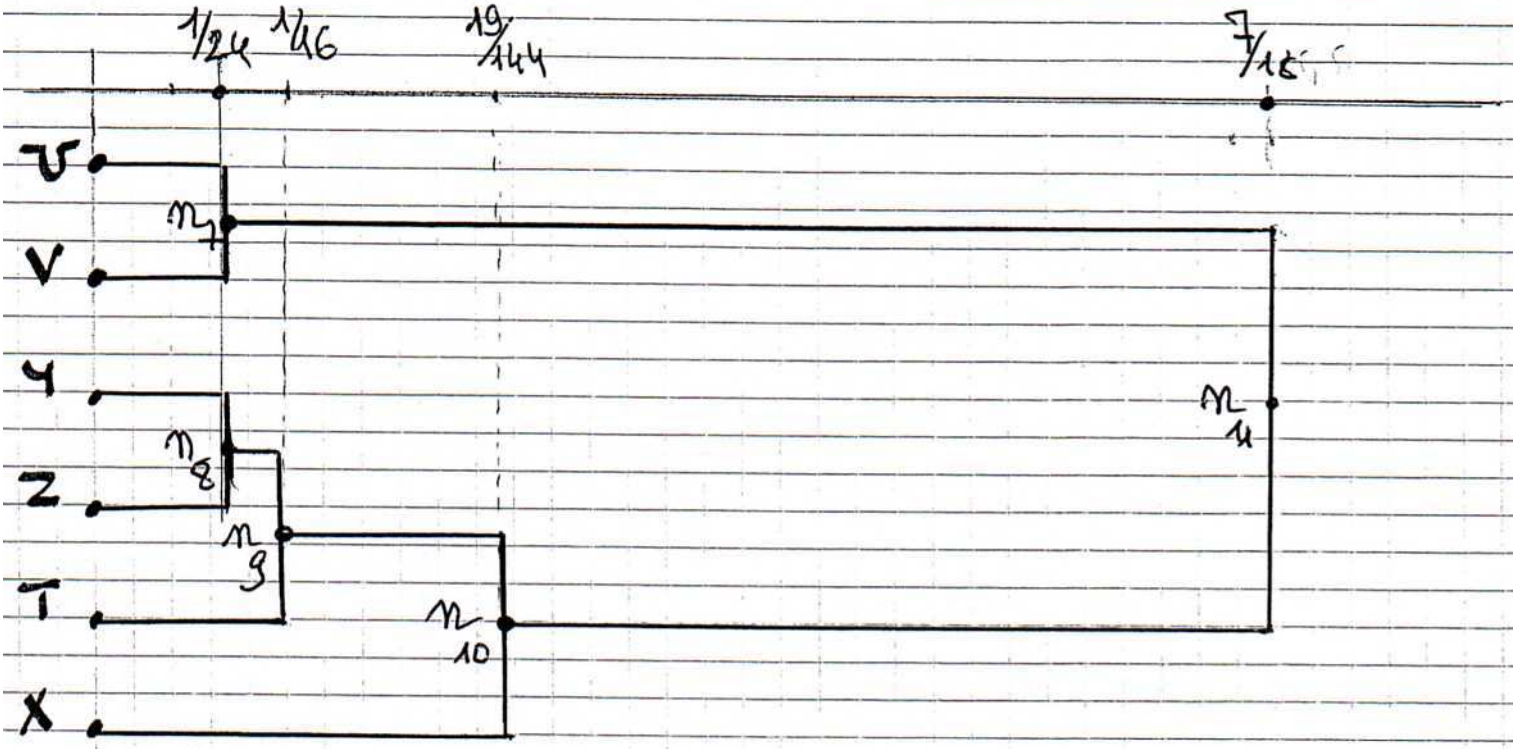
On agrège X et $n_9$ $\rightsquigarrow$ $n_{10}$ (barycentre de X et $n_9$)

| $d^2$ | $n_{10}$ | $n_7$ |
|---|---|---|
| $n_{10}$ | 0 | 56/27 |
| $n_7$ | " | 0 |

| $\delta$ | $n_{10}$ | $n_7$ |
|---|---|---|
| $n_{10}$ | 0 | (7/18) |
| $n_7$ | " | 0 |

On agrège enfin $n_{10}$, $n_7$ $\rightsquigarrow$ $n_{11}$ (barycentre de $n_{10}$, $n_7$)
C'est le barycentre du nuage.

En respectant les niveaux des nœuds, on a donc le dendrogramme ci-dessous.

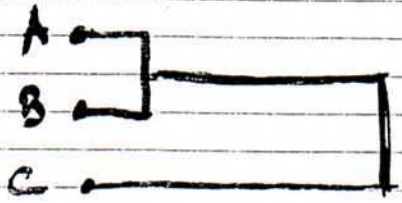



On vérifie au passage que la somme des niveaux de nœuds :

$$v(n_7)+v(n_8)+v(n_9)+v(n_{10})+v(n_{11})$$
$$= \frac{1}{24}+\frac{1}{24}+\frac{1}{16}+\frac{19}{144}+\frac{7}{18}=\frac{2}{3}$$ c'est à dire

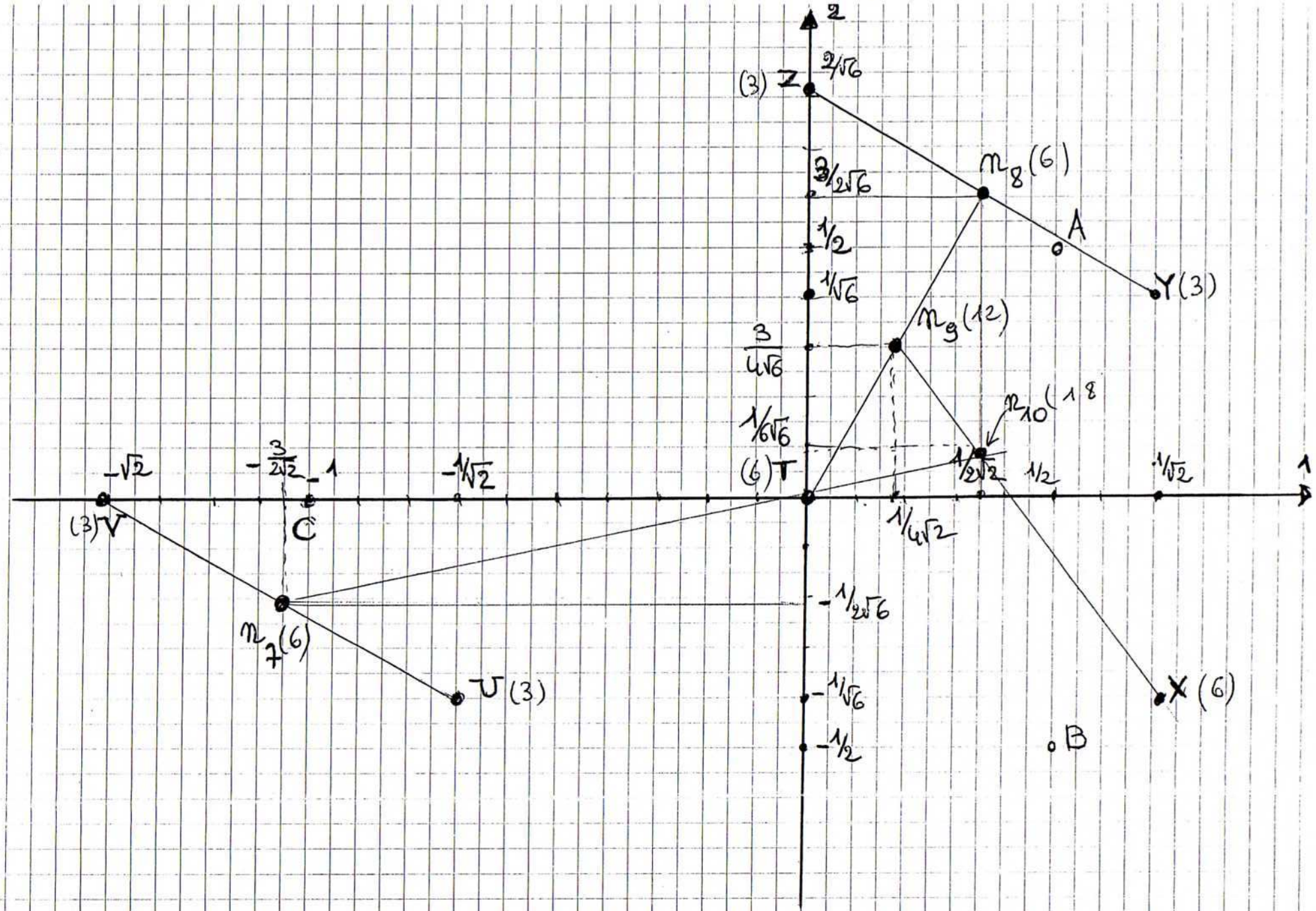
l'inertie du nuage, c'est à dire $d_1+d_2=\frac{1}{2}+\frac{1}{6}$

Pour les lignes le dendrogramme




C correspond à (V,V) (AB) sera associé à l'ensemble X42T

feuille jointe



# Analyse des données 2ème session 30 août 2007

**Aucun document autre qu'un polycopié de cours n'est autorisé, aucune calculette, aucun téléphone.**

## QUESTION DE COURS :

Expliquer le principe de base de la méthode des classifications
Pourquoi définit-on un indice de niveau ?
Cette méthode apporte t'elle des informations supplémentaires par rapport à l'A.F.C. ou l'A.C.P. ?

## PROBLEME :

Soit le tableau de données ci-contre :

| | X | Y | Z |
|---|---|---|---|
| A | 0 | 0 | 3 |
| B | 0 | 1 | 2 |
| C | 1 | 1 | 1 |
| D | 1 | 1 | 1 |
| E | 2 | 0 | 1 |
| F | 2 | 1 | 0 |
| G | 1 | 2 | 0 |
| H | 1 | 2 | 0 |

### Première partie : Analyse en Composantes Principales :

1) Exhiber le tableau des données centrées réduites (remarque : étant en ACP, on ne rassemblera pas les individus ayant même ligne de données)
2) En déduire la matrice à diagonaliser.
3) Calculer les valeurs propres et les vecteurs propres de cette dernière.
4) En déduire les coordonnées des représentants des lignes et des colonnes.
5) Dessiner sur un même graphique les individus et les variables de l'unique plan factoriel (une des trois valeurs propres étant nulle).

### Deuxième partie : Classification Ascendante Hiérarchique :

1) En vue de faire la CAH des lignes de ce tableau, on dressera le tableau des distances au carré entre individus (nous sommes en ACP, on prend donc la distance euclidienne) ; remarquons que l'on peut trouver aisément ces distances à partir du graphique de la première partie et agréger dès le début C&D, d'une part et G&H d'autre part.
2) En déduire le tableau des indices de niveau, les individus ayant tous le même poids, par exemple 1 (et donc 2 pour C&D et pour G&H)
3) En déduire les premiers éléments à agréger, puis continuer l'algorithme.
4) Dessiner le dendrogramme correspondant en tenant compte des indices de niveau.
5) Reporter cette classification sur le graphique du plan factoriel de la première partie.
6) Cette classification est-elle bien en cohérence avec l'ACP précédente ?

INT: AD. Août 2007

| | X | Y | Z | |
|---|---|---|---|---|
| A | 0 | 0 | 3 | 3 |
| B | 0 | 1 | 2 | 3 |
| C | 1 | 1 | 1 | 3 |
| D | 1 | 1 | 1 | 3 |
| E | 2 | 0 | 1 | 3 |
| F | 2 | 1 | 0 | 3 |
| G | 1 | 2 | 0 | 3 |
| H | 1 | 2 | 0 | 3 |
| | 8 | 8 | 8 | |

## Première partie: A.C.P.

### 1) Données centrées réduites:

On a $\bar{X} = \bar{Y} = \bar{Z} = 1$

$\sigma_X^2 = \sigma_Y^2 = \frac{1}{2}$ , $\sigma_Z^2 = 1$

La matrice $X = \left\{ x_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \right\}$

s'écrit donc:

$$X = \begin{pmatrix} -\sqrt{2} & -\sqrt{2} & 2 \\ -\sqrt{2} & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \sqrt{2} & -\sqrt{2} & 0 \\ \sqrt{2} & 0 & -1 \\ 0 & \sqrt{2} & -1 \\ 0 & \sqrt{2} & -1 \end{pmatrix}$$

et donc:

### 2) Matrice à diagonaliser:

$$^tXX = \begin{pmatrix} 8 & 0 & -4\sqrt{2} \\ 0 & 8 & -4\sqrt{2} \\ -4\sqrt{2} & -4\sqrt{2} & 8 \end{pmatrix}$$

### 3) Les valeurs propres et données par:

$\det(^tXX - \lambda I) = 0$

$\pi$: $^tXX - \lambda I$ peut s'écrire $+4\sqrt{2} \begin{pmatrix} \gamma & 0 & -1 \\ 0 & \gamma & -1 \\ -1 & -1 & \gamma \end{pmatrix}$

avec: $\gamma = \sqrt{2} - \frac{\lambda}{4\sqrt{2}}$

ou. $\begin{vmatrix} 4 & 0 & -1 \\ 0 & 4 & -1 \\ -1 & -1 & 4 \end{vmatrix} = 0 \Rightarrow 4^3 - 24 = 0 \Rightarrow 4 = 0, 4 = \pm\sqrt{2}$

On a donc: puisque $\Delta = 4\sqrt{2}(\sqrt{2} - 4)$

$\delta_1 = 16$, $\delta_2 = 8$, $\delta_3 = 0$

(l'A.C.P. sera réduite ici à un plan puique $\delta_3 = 0$

Résolvant ${}^t X X u = \delta u$ :

On en déduit: à $\delta_1 = 16$ $\rightsquigarrow$ ${}^1u \begin{vmatrix} 1/2 \\ 1/2 \\ -1/\sqrt{2} \end{vmatrix}$

à $\delta_2 = 8$ $\rightsquigarrow$ ${}^2u \begin{vmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{vmatrix}$

4) <u>Les coordonnés des individus</u> (ou représentants des lignes) sont donnés par:

$$F_\alpha(i) = \sum_{j=1}^{3} x_{ij} \, {}^\alpha u_j$$

Matriciellement, on obtiendra les coordonnés de l'ensemble des points en faisant le produit:

$$F = XU \quad \text{avec} \quad U = \begin{pmatrix} 1/2 & 1/\sqrt{2} \\ 1/2 & -1/\sqrt{2} \\ -1/\sqrt{2} & 0 \end{pmatrix} \text{ ou encore } \frac{1}{\sqrt{2}} \begin{pmatrix} 1/\sqrt{2} & 1 \\ 1/\sqrt{2} & -1 \\ -1 & 0 \end{pmatrix}$$

en mettant $\sqrt{2}$ en facteur de $X$ : on peut encore écrire

~ 3 x

$$F = X.U = \sqrt{2}\,\frac{1}{\sqrt{2}} \begin{pmatrix} -1 & -1 & \sqrt{2} \\ -1 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1/\sqrt{2} \\ 0 & 1 & -1/\sqrt{2} \\ 0 & 1 & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1 \\ 1/\sqrt{2} & -1 \\ -1 & 0 \end{pmatrix} = \begin{matrix} \text{axe 1} & \text{axe 2} \\ \begin{pmatrix} -2\sqrt{2} & 0 \\ -\sqrt{2} & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 2 \\ \sqrt{2} & 1 \\ \sqrt{2} & -1 \\ \sqrt{2} & -1 \end{pmatrix} \end{matrix} \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \end{matrix}$$

et $G_{ij} = \sqrt{\frac{d_j}{n}}\, u_{ij}$ c.à.d. matriciellement :

puisque $d_1 = 16$, $d_2 = 8$, $n = 8$

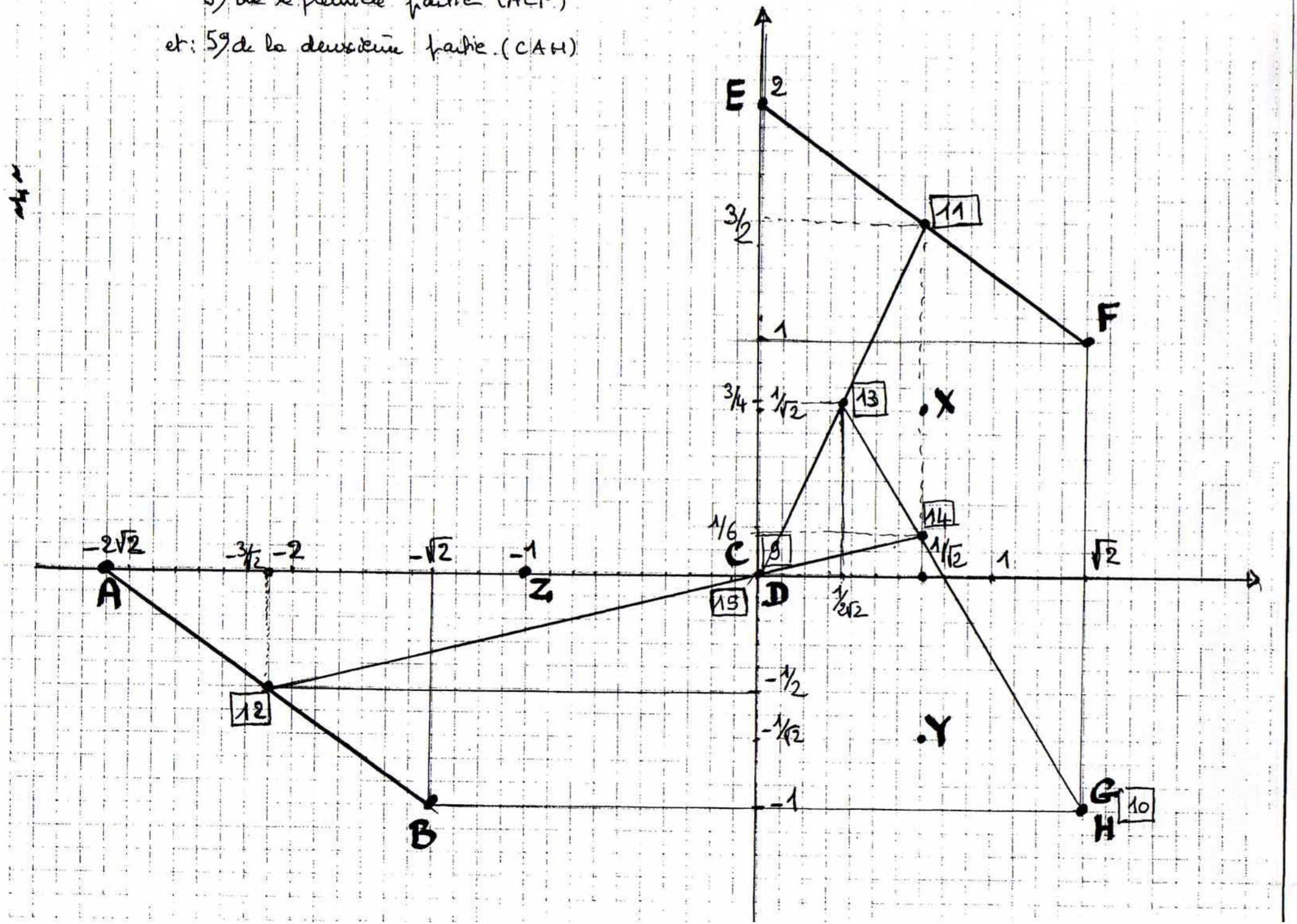
$$G = U \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{pmatrix} = \begin{matrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \\ -1 & 0 \end{pmatrix} \begin{matrix} x \\ y \\ z \end{matrix} \end{matrix} \quad \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \underline{\text{coordonnées des variables}}$$

axe 1 axe 2

On retrouve bien que X, Y et Z sont sur une sphère unité (ici réduite au cercle bien sûr).

5) de la première partie (ACP)

et: 5) de la deuxième partie. (CAH)



## Deuxième partie: C.A.H:

1) Distances entre individus. (ayant regroupé C,D et G H qui peuvent être considérés ici comme un individu de poids 2 chacun).

| $d^2$ | A | B | CD | E | F | GH |
|---|---|---|---|---|---|---|
| A | 0 | 3 | 8 | 12 | 19 | 19 |
| B | | 0 | 3 | 11 | 12 | 8 |
| CD | | | 0 | 4 | 3 | 3 |
| E | | | | 0 | 3 | 11 |
| F | | | | | 0 | 4 |
| GH | | | | | | 0 |

L'indice de distance est donné par:

$$\delta(i,i') = \frac{m_i m_{i'}}{m_i + m_{i'}} d^2(i,i')$$

ici l'on a tous commencé:

$$m_A = m_B = m_E = m_F = 1$$

$$m_{CD} = m_{GH} = 2 \quad \text{et donc le}$$

premier tableau d'indice de distance :

2)

| δ | A | B | 9 (CD) | E | F | 10 (GH) |
|---|---|---|---|---|---|---|
| A | 0 | 3/2 | 16/3 | 6 | 19/2 | 38/3 |
| B | | 0 | 2 | 11/2 | 6 | 16/3 |
| CD | | | 0 | 8/3 | 2 | 3 |
| E | | | | 0 | 3/2 | 22/3 |
| F | | | | | 0 | 8/3 |
| GH | | | | | | 0 |

3)

On peut considérer que le noeud 9 était CD, le noeud 10 était GH avec $\nu(9) = \nu(10) = 0$ on notera donc 11 le noeud obtenu en agrégeant E et F (on aurait pu commencer par : A et B) avec $\nu(11) = 3/2$.

4) On a donc les tableaux suivants permettant de continuer l'algorithme :

| $d^2$ | A | B | 9 (CD) | 11 (EF) | 10 (GH) |
|---|---|---|---|---|---|
| A | 0 | 3 | 8 | 59/4 | 19 |
| B | | 0 | 3 | 43/4 | 8 |
| 9 (CD) | | | 0 | 11/4 | 3 |
| 11 (EF) | | | | 0 | 27/4 |
| 10 (GH) | | | | | 0 |

EF ou 11 a pour coordonnée $(1/\sqrt{2}, 3/2)$

| δ | A | B | 9 (CD) | 11 (EF) | 10 (GH) |
|---|---|---|---|---|---|
| A | 0 | 3/2 | 16/3 | 59/6 | 38/3 |
| B | | 0 | 2 | 43/6 | 16/3 |
| 9 (CD) | | | 0 | 11/4 | 3 |
| 11 (EF) | | | | 0 | 11/4 |
| 10 (GH) | | | | | 0 |

⇐ On agrège donc A, B comme le laissait prévoir 3).

Le noeud 12 est donc l'agrégation de A et B avec $\nu(12) = 3/2$. Il vient donc : les coordonnées du noeud 12 étant $(-3/\sqrt{2}, -1/2)$.

| (δ) $d^2$ | 12 (AB) | 9 (CD) | 11 (EF) | 10 (GH) |
|---|---|---|---|---|
| 12 (AB) | 0 | 19/4 | 12 | 51/4 |
| 9 (CD) | | 0 | (11/4) | 3 |
| 11 (EF) | | | 0 | 27/4 |
| 10 (GH) | | | | 0 |

Remarque : tous les poids ici sont égaux à 2 et donc : $\frac{m_i m_{i'}}{m_i + m_{i'}} = 1$ : le tableau de δ est identique à celui de $d^2$

Le noeud 13 sera donc l'agrégation de 9 et 11 avec $\nu(13) = 11/4$ coordonnées $\left(\frac{1}{2\sqrt{2}}, \frac{3}{4}\right)$

On a donc le tableau suivant :

| $d^2$ | 12 | 13 | 10 |
|---|---|---|---|
| 12 (AB) | 0 | 123/16 | 51/4 |
| 13 (9,11) | | 0 | 67/16 |
| 10 (GH) | | | 0 |

et donc le tableau des indices :

| δ | 12 | 13 | 10 |
|---|---|---|---|
| 12 | 0 | 41/4 | 51/4 |
| 13 | | 0 | (67/12) |
| 10 | | | 0 |

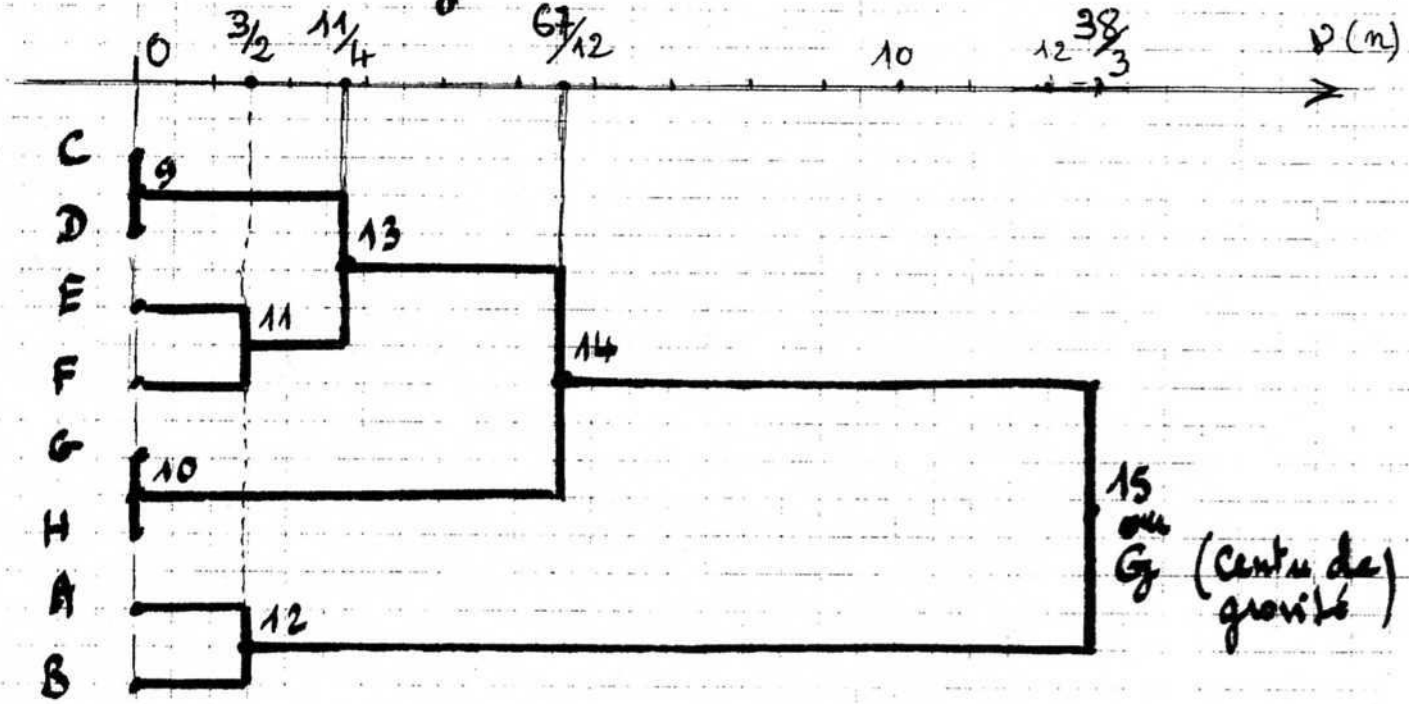
Le noeud 14 suivant sera donc l'agrégation de 13 et 10 ; compte-tenu des poids respectifs, il aura pour coordonnées $\left(\frac{1}{\sqrt{2}}, \frac{1}{6}\right)$ et $\nu(14) = \frac{67}{12}$

Il reste à agréger 12 et 14 ; on remarque que le barycentre est bien à l'origine.

De plus $d^2(12,14) = \frac{76}{9}$ et compte-tenu des poids respectifs 2 et 6 on a $\delta = \frac{2.6}{2+6} \cdot \frac{76}{9} = \frac{38}{3}$

On a donc : $P(15) = \frac{38}{3}$.

4) D'où le dendrogramme :




On vérifie, au passage, que la somme des niveaux des nœuds est égale à la somme des valeurs propres c.à.d. à l'inertie totale.

En effet $0 + \frac{3}{2} + \frac{3}{2} + \frac{11}{4} + \frac{67}{12} + \frac{38}{3} = \frac{288}{12} = 24$

et $d_1 + d_2 = 16 + 8 = 24$ c.q.f.d.

6) La cohérence avec l'A.C.P. selon que les individus sont concernés par x, y ou z apparaît bien.