

# TD: ACP

## Exercice I:

• Analyse ACP réalisée sur 50 avions.

• pour chaque avion on a étudié 10 variables.

1. Ce sont les variables représentées par des points proches du cercle de corrélation et proche de l'axe 1 : donc ici : V<sub>6</sub>, V<sub>7</sub> et V<sub>5</sub>.

2. Ce sont les variables proche de l'origine alors elles sont peu corrélées avec les axes. V<sub>2</sub> et V<sub>3</sub> sont peu corrélés avec les axes.

3. Fortement corrélées positivement entre elles ; donc coeff de corrélat est proche de 1 et positif  
⇒ ces angles entre ces deux variables est proche de 1 et positif - donc angle aigu entre ces deux variables. [cor  $\cos(x,y) = \rho(x,y)$ ]

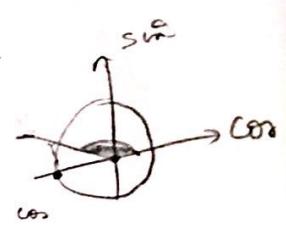
G<sub>1</sub>: V<sub>5</sub> et V<sub>7</sub>

G<sub>2</sub>: V<sub>4</sub>, V<sub>8</sub> et V<sub>10</sub>

G<sub>3</sub>: V<sub>1</sub> et V<sub>9</sub>.

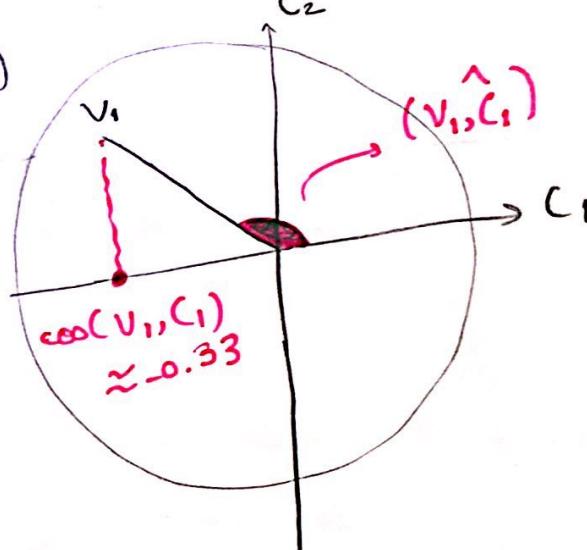
4. Variables peu corrélées entre elles : forment entre elles un angle proche de  $\frac{\pi}{2}$ . (plus de  $\frac{\pi}{2}$  ou moins de  $\frac{\pi}{2}$ )  
V<sub>1</sub> et V<sub>6</sub>; V<sub>1</sub> et V<sub>5</sub>; V<sub>9</sub> et V<sub>6</sub>  
V<sub>5</sub> et V<sub>9</sub>; V<sub>4</sub> et V<sub>7</sub>; V<sub>1</sub> et V<sub>8</sub>

①

5) Fortement corélés négativement avec  $V_4 (> \frac{\pi}{2})$   
 → corrélation négative forme un angle obtuse avec  $V_4$  (proche de  $\pi$ ). 

Ici  $V_1$  et  $V_3$ .

6) coefficient de corrélation entre  $V_1$  et l'axe  $C_1$   
 égale au cosinus de l'angle entre  $V_1$  et  $C_1$   
 $g_r(V_1, C_1) = \cos(\hat{V_1}, \hat{C_1})$   
 $= -0.33$



Variable  $V \approx$   
 7) Corrélation presque 1 avec  $C_2$ .  
 donc  $C_2$  et la variable  $V$  sont très proches  
 $\Rightarrow V = V_4$  ou  $V_3$ .

8) Corrélation entre  $V_3$  et  $V_7$ ?  
 On peut dire que la variable  $V_3$  est déjà mal représentée.

9) Les deux composantes principales ( $C_1$ ) et ( $C_2$ ) sont orthogonales (angle  $C_1, C_2 = \frac{\pi}{2}$ ): donc coeff est nul. ②  
 $\rightarrow \cos \frac{\pi}{2} = 0$

## Exercice 2

1<sup>e</sup>. Vrai.

plus les variables sont corrélées entre elles, plus il est facile de les résumer par un petit nombre de variables synthétiques (les CP) et donc plus le pourcentage d'inertie porté par les premières CP est importante.

2<sup>e</sup>. Vrai.

Les individus ont un poids (un importance), les individus très éloignés du centre de gravité du nuage contribuent à une part importante de la variabilité (l'inertie).

3<sup>e</sup>. Vrai.

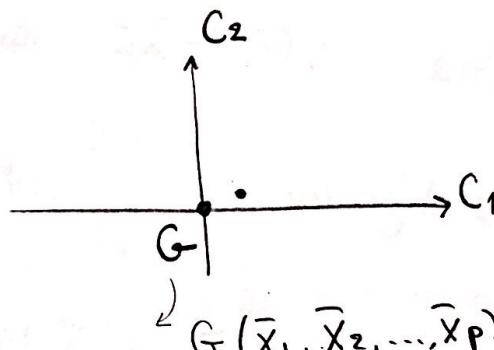
La variance sur le premier axe  $\text{Var}(C_1) = \lambda_1$   
et  $\text{Var}(C_2) = \lambda_2$ ; et on sait que  $\lambda_1 > \lambda_2$ .

4<sup>e</sup>. Faux.

Pour que elles soient très corrélées, il faut qu'elles soient déjà bien représentées.

5<sup>e</sup>. Faux.

un individu proche du G, c'est à dire il a des valeurs proches au G.



①

## Exercice 3.

### ① Matrice de corrélation:

Toutes les corrélations sont positives, ce qui signifie que toutes les variables varient dans le même sens.

En particulier, Math et physique sont très corrélées (~0.98) ainsi que les deux variables "ANGL" et "FRAN"; alors que "Math" et "français" sont très peu corrélés (~0.23).

ACP: on remarque alors deux groupes de variables fortement corrélées, cela laisse penser que l'information pourra être représentée sur les premiers axes de l'ACP.

### ② Résultats sur l'inertie:

a)  $\text{Var}(C_i) = \lambda_i$

b) On a:  $V(C_i) = \lambda_i$  alors: 
$$\sum_{i=1}^p V(C_i) = \sum_{i=1}^p \lambda_i = I$$

on sait déjà que:

$$\sum_{i=1}^p V(X_i) = I$$

alors: 
$$\sum_{i=1}^p V(X_i) = \sum_{i=1}^p V(C_i) = I$$

Le nuage de point en dimension (4) est toujours le même et sa dispersion globale n'a pas changé, donc soit dans le plan initial soit dans le nouveau, toujours on a même  $I$ .

9

### ③ Résultats sur les variables:

- Avec le tableau "valeurs propres et variance expliquée" on peut remarquer que le premier composant principal contient 70% de l'information, et le deuxième contient 30% alors on retient juste ces 2 composantes. ( $C_3$ ) et ( $C_4$ ) ne seront ni utilisés ni interprétés.
- Avec le tableau "Corrélation variables-composantes"; on remarque que  $C_1$  est corrélé positivement et fortement avec les 4 variables car toutes les corrélat° sont positives, alors plus l'élève a des bonnes notes dans ces 4 disciplines, plus il a un score (abscisse) élevé sur l'axe  $C_1$ .  
Donc, on peut dire que  $C_1$  représente en quelques sortes les résultats globaux des élèves.
- Pour l'axe  $C_2$ : Il oppose d'un part le français et l'anglais (corrél° positives 0,66 et 0,41) et d'autre part le math et le physique (corrél° négatives -0,58 et -0,43)  
Donc, il s'agit d'un axe d'opposition entre discipline scientifique et littéraire.  
C.-à-d: { plus note en littérature élève plus l'ordonnée est élevée  
{ plus note en scientifq élève plus l'ordonnée est moins élevée

③

- La corrélation entre les variables d'un part, ou entre eux et les axes  $(C_1), (C_2)$  d'autre part peuvent s'interpréter en regardant les angles entre eux, plus l'angle est aigu, plus le  $\cos$  est élevé alors la corrélation aussi ...

#### ④ Résultats sur les individus:

- ④ Résultats sur les axes

  - poids = pondération =  $\frac{1}{n} = \frac{1}{g} = 0,11$  > C'est à dire tous les individus ont même importance.  
nb d'individus
  - FACT 1:** donne les coordonnées de l'individu sur l'axe  $C_1$  (abscise) et  $C_2$  (ordonnée).
  - FACT 2:** " " " "
  - CONTG:** Donne la contribution générale de l'individu (dans  $\mathbb{R}^P$ ) à la construction du nouveau plan.
  - CONT1:** contribution de l'individu en % à la construction de l'axe 1
  - CONT2:** " " " "
  - cosCa1:** cosinus corré, donne la qualité de représentation de l'individu sur l'axe 1.
  - cosCa2:** " " " "
  - sur l'axe 2.

$$\cos C_1 \cdot 1 + 0.5 \cos C_2 = 1 \quad \text{c.-à-d l'individu est}$$

bien représenté dans le nouveau plan  $(C_1, C_2)$ .

→ Interprétation de tableau "Coordonnées des individus contributifs" en choisit l'individu "Jean":

$(-8,61)$  et  $(-1,41)$  sont les coordonnées de Jean dans le nouveau plan  $(C_1, C_2)$ .

•  $\text{CONTG} = 20,99$  c.-à-d Jean contribue à 20,99% à la construction du nouveau plan  $(C_1, C_2)$ .

•  $\text{CONT1} = 29,19$ , alors Jean contribue à 29,19% à la construction de l'axe  $C_1$ .

•  $\text{CONT2} = 1,83$ , alors Jean contribue à 1,83% à la construction de l'axe  $C_2$ .

$$\text{CONT 1} = \frac{\text{Poids}(i) \text{ (FACT 1)}}{\lambda_1} \times 100$$

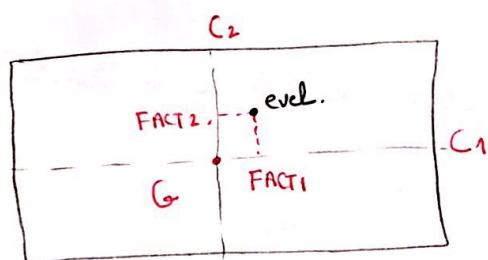
$$\text{pour Jean, } \text{CONT1} = \frac{\frac{1}{9} \cdot (-8,61)}{28,93} \times 100 = 29,19\%$$

$$\text{CONT 2} = \frac{\text{Poids}(i) \cdot (\text{FACT 2})}{\lambda_2} \times 100$$

$$\text{pour Jean, } \text{CONT2} = \frac{\frac{1}{9} \cdot (-1,41)}{12,03} \times 100 = 1,83\%. \quad (5)$$

- cosca 1 (Jean) = 0,97
  - cosca 2 (Jean) = 0,03
- $0,97 + 0,03 = 1$
- Donc, Jean est bien représenté dans le nouveau plan  $(C_1, C_2)$ .

**• Interprétation du graphe est individualisée :**



**Sur l'axe  $(C_2)$ :**

- plus les individus sont en haut dans le graphe, plus leurs ordonnées (ordonnée sur l'axe  $C_2$ ) sont élevées; alors ils ont des bonnes notes en littérature (corrélation positive avec cet axe).
- au contraire, les individus en bas ont des notes meilleures en scientifiques (corrélation négative avec cet axe).

Exp:  $\rightarrow$  Pierre le plus haut a des bonnes notes en littérature (14 et 11,5) contre des mauvaises notes en scientifique (5,5 et 7).

$\rightarrow$  André le plus bas est bien en scientifiques (11 et 10) mais pas en littérature (5,5 et 7).

De plus: « Moni » et « Alan » ont un ordonnées proche de l'origine, alors, « chacun d'eux a les mêmes notes en ⑥ scientifiques et en littéraires. (voir leur notes).

\* Sur l'axe ( $c_1$ ): (qui représente la moyenne générale).  
plus la moyenne de l'étudiant est élevée plus son  
coordonnée sur l'axe  $c_1$  (absisse) va être élevée.

Exp: "moni" a l'absisse le plus élevée, alors elle a la  
moyenne générale la plus élevée: (qui est 15)  
"jean" a l'absisse le moins élevé, donc la moyenne  
Le moins élevé (qui est 5,6).

the end

F1)