

TD: ACP.

Exercice I:

- Analyse ACP réalisé 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_6, V_7 et V_5 .

2. Ce sont les variables proche de l'origine alors ils sont peu corrélés avec les axes. V_2 et V_3

3. Fortement corréliées positivement entre elles ; donc coeff de corrélat est proche de 1 et positive $\Rightarrow$ cos angle entre ces deux variables est proche de 1 et positive - donc angle aigu entre ces deux variables. [$\cos \cos(x, y) = r(x, y)$]

G1: V_5 et V_7

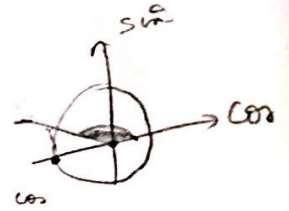
G2: V_4, V_8 et V_{10}

G3: V_1 et V_9 .

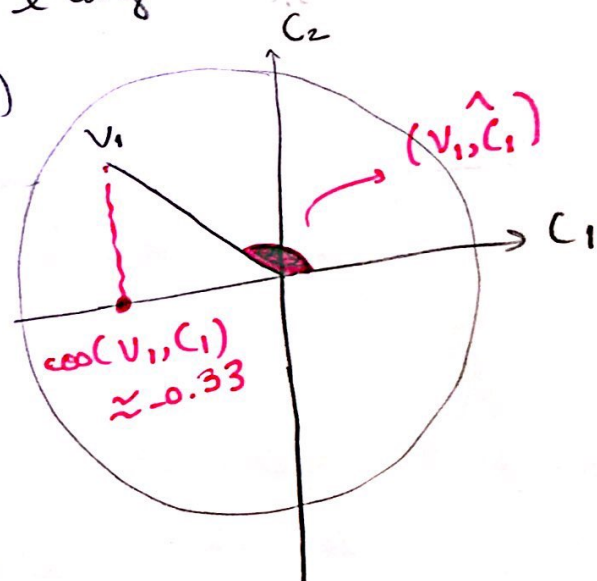
4. Variables peu corrélées entre elles : forment entre elle un angle proche de $\frac{\pi}{2}$. (plus de $\frac{\pi}{2}$ ou moins de $\frac{\pi}{2}$)
 V_5 et V_9 ; V_4 et V_7 ; V_1 et V_6 ; V_1 et V_5 ; V_9 et V_6

①

5) Fortement corréles négativement avec V_4 ($> \frac{\pi}{2}$)
 $\rightarrow$ cos négatif $\rightarrow$ forme un angle obtus avec V_4 (proche de π).
 Ici V_1 et V_9 .



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



Variable $V \approx$

7) Corrélat vaut presque 1 avec C_2 .
 donc C_2 et la variable V sont très proches
 $\Rightarrow V = V_4$ ou V_9 .

8) Corrélat entre V_3 et V_7 ??

On peut même 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 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 corrélés (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 corrélés, 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: $\text{Var}(C_i) = \lambda_i$ alors:

$$\sum_{i=1}^p \text{Var}(C_i) = \sum_{i=1}^p \lambda_i = I$$

on sait déjà que:

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

alors:

$$\sum_{i=1}^p \text{Var}(X_i) = \sum_{i=1}^p \text{Var}(C_i) = I$$

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 composante principal contient 70% de l'information, et le deuxième contient 30%, alors on retient juste ces 2 composantes. (C3) et (C4) ne seront ni utilisés ni interprétés.

Avec le tableau "Corrélation variables-composantes": on remarque que C1 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 C1.
Donc, on peut dire que C1 représente en qdq sortes les résultats globaux des élèves.

Pour l'axe C2: 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.

C1 a-d:
| plus note en littéraire élevée plus l'ordonnée est élevée.
| plus note en Scientif° élevée plus l'ordonnée est moins élevée

③

5.

• poids = pondération = $\frac{1}{n} = \frac{1}{9} = 0,11$: c-à-d tous les individus ont m même importance.

FACT 2: " " "

CONT 1 : contribution de l'individu en % à la construction de l'axe 1

CONT 2:

CONT 2: "
CosCa1: cosinus corré, donne la qualité de représentat de l'individu
sur l'axe 1.

Costa 2:

sur l'axe.

④

• $\text{cosca 1 (Jean)} = 0,97$

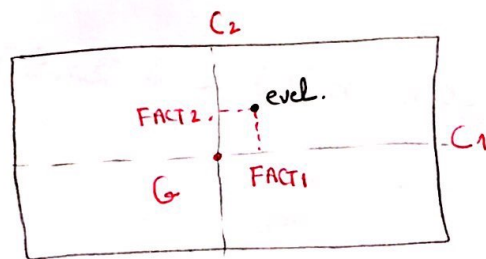
• $\text{cosca 2 (Jean)} = 0,03$

$$> 0,97 + 0,03 = 1$$

Donc, Jean est bien représenté dans le,

nouveau plan (C_1, C_2) .

• **Interprétation du graphe est indiquée :**



→ **Sur l'axe (C_2) :**

• plus les individus sont en haut dans le graphe, plus leurs ordonnées (ordonnées sur l'axe C_2) sont élevées ; alors ils ont des bonnes notes en littéraire (car corrélat positif avec cet axe).

au contraire, les individus en bas ont des notes meilleures en scientifiques (car corrélat négatif avec cet axe).

Exp: → Pierre le plus haut a des bonnes notes en littéraire (14 et 11,5) contre des mauvais notes en scientifique (5,5 et 7).

→ André le plus bas est bien en scientifiques (11 et 10) mais pas en littéraire (5,5 et 7).

De plus : « Moni » et « Alan » ont un ordonnées proche de 0. L'origine, alors, "chacun d'eux a les mêmes notes en scientifiques et en littéraire. (voir leur notes). (0-0)

* 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 (absurx) va être élevée.

Exp: "mon" a l'absurx le plus élevée, alors elle a la
moyenne générale le plus élevée: (qui est 15)
"Jean" a l'absurx le moins élevée, donc la moyenne
le moins élevée (qui est 5.6).

the end