

17/20

16

Université Claude Bernard - Lyon 1 I.S.F.A.

M1

Année 2009-2010

Contrôle Continu d'Analyse des Données
Mercredi 21 Octobre 2009
Durée : 1h30

A.B. Dufour

Tous les documents sont autorisés.

Nom, Prénom MANAS Adrien

1 Etudiants et bien-être

Six étudiants ayant passé un baccalauréat "S" ou "ES" réalisent des études de mathématiques appliquées. A la fin du premier semestre, on leur demande de noter (sur une échelle de 0 à 5) la façon dont ils se sentent intégrés et soutenus à l'université.

baccalauréat	intégration	soutien
S	3	2
ES	2	3
ES	4	2
S	4	2
ES	2	4
S	3	3

Les responsables de l'enquête considèrent que les étudiants venant de la filière "ES" sont moins bien préparés à ce type d'étude. C'est pourquoi il est accordé un poids une fois et demi plus important à la réponse d'un étudiant provenant du bac "ES".

Donner, pour chaque question posée, l'expression de \mathbb{P} et le résultat obtenu.

$$\sum p = 1$$

$$3p + 3 \times 2p = 1$$

$$3(1-p) + 2p = 1 \quad \text{BBE}$$

$$p = \frac{1}{3} =$$

A.B. Dufour



1. Construire le vecteur de pondération associé aux individus.

Réponse :

$$P \leftarrow c\left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}\right)$$

on demande
1.5 fois plus et non 2

$$0,5 \begin{matrix} N \\ / \\ 1,5 \end{matrix}$$

$$P = c(1; 1.5; 1.5; 1; 1.5; 1)$$

$$P = \begin{pmatrix} 1 & 1.5 & 1.5 & 1 & 1.5 & 1 \\ 0.5 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix}$$

2. Calculer la moyenne générale de l'intégration ainsi que sa valeur numérique (deux chiffres après la virgule).

Réponse :

$$\text{intégration} \leftarrow c(3, 2, 4, 4, 2, 3) \\ \text{moy} \leftarrow \text{sum}(p * \text{intégration}) \\ \text{moy}$$

$$\Rightarrow \text{moy} = 2,89 \quad 2,93$$

3. Calculer l'angle entre les deux vecteurs intégration et soutien.

Réponse :

$$\text{mame} \leftarrow \text{function}(x, \text{pds}) \quad \text{sout} \leftarrow c(2, 3, 2, 2, 4, 3) \\ \text{mame} \leftarrow \text{sum}(\text{pds} * x * x) \\ \text{mame} (\text{intégration}, p) \quad \text{prodscal} \leftarrow \text{function}(x, y, \text{pds}) \quad \text{sum}(\text{pds} * x * y) \\ 3,02 \quad \text{cosangle} \leftarrow \text{prodscal}(\text{intégration}, \text{souten}, \text{pds}) / \\ \text{mame} (\text{soutien}, p) \quad (\text{mame} (\text{intégration}, \text{pds}) * \text{mame} (x, y) / \text{mame} (\text{soutien}, \text{pds})) \\ 2,89 \\ \text{pascal} \leftarrow \text{E}(x) \% * \% \cdot \text{intégration} \quad \text{cosa} = \frac{\text{cosangle}}{\|\text{x}\| \cdot \|\text{y}\|}$$

par simple multiplication / division des résultats obtenus

l'astuce cosa = 5,16
La valeur est fausse
puisque cosa < 1 nécessairement

4. Interpréter, d'un point de vue statistique, l'angle entre ces deux vecteurs.

Réponse :

Le fait que cosa se rapproche de 1 signifierait que les deux variables sont liées. Inversement si cosa est proche de 0, les deux variables sont indépendantes.

$$0,5 \begin{matrix} N \\ / \\ 1,5 \end{matrix}$$

$$2 + (0,5)$$

2 Religion et statut de l'embryon

Dans l'étude analysée lors des séances de TD, on s'est intéressé au statut de l'embryon dans différents pays européens : Danemark (DA), Italie (IT), Espagne (ES), France (FR), Pologne (PO), Royaume-Uni (RU), Pays-Bas (PB), Allemagne (AL) et Autriche (AU).

Le statut de l'embryon est réparti en quatre catégories : ensemble de cellules (CC), cellules allant donner naissance à un être humain (CCHB), être humain provenant d'un ensemble de cellules (HBCC) et être humain (HB). Les non réponses ne sont pas incluses dans l'analyse. On s'intéresse aux 10343 réponses données par les enquêtés pratiquant une religion : catholicisme (C), protestantisme (P), anglicanisme (AN), évangelisme (EV).

Une analyse des correspondances est réalisée sur la table de contingence croisant pays et statut de l'embryon.

- (a) Donner, en justifiant la réponse, le nombre total d'axes principaux.

Réponse :

N'ayant pas le graphe des valeurs propres, je ne peux pas savoir le nombre d'axes principaux à nettoyer

I : nombre de modalités de la variable V₁

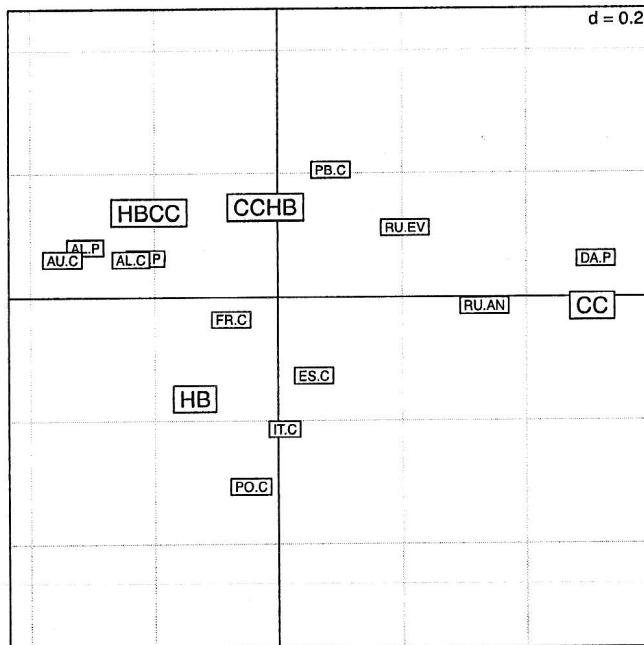
J : nombre de modalités de la variable V₂

*vous n'avez pas
compris la
question*

$$0 \begin{matrix} N \\ / \\ 1,1 \end{matrix}$$

$$lf(I-1, J-1) = (9-1, 4-1) = 3$$

- (b) Commenter la représentation donnée par le premier plan factoriel.



Réponse :

La projection sur l'axe horizontal fait apparaître l'opposition de point de vue entre les français protestants qui considèrent l'émigration comme un ensemble de cellules et les autres catholiques et les allemands catholiques et protestants qui considèrent l'émigration un être humain provenant d'un ensemble de cellules.

3

- | La projection sur l'axe vertical donne peu d'informations car les projections sont proches de l'origine du repère pour la plupart des groupes. On peut toutefois remarquer que les italiens catholiques, les polonais catholiques et les espagnols catholiques considèrent l'émigration comme un être humain.
- (c) La valeur du Chi-Deux observée est 907.45. Sachant que les deux premières valeurs propres valent respectivement 0.06425 et 0.01876, donner le pourcentage d'inertie expliquée par les deux premiers facteurs de l'analyse réalisée.

Réponse :

$$\text{On sait que } I_T = \frac{\chi^2}{n} \text{ et } \chi^2 = 907,45 \\ n = 10343$$

$$\Rightarrow I_T = 0,08773$$

$$\text{donc } \% = \frac{0,06425 + 0,01876}{0,08773} = 0,946$$

Les deux premiers facteurs de l'analyse réalisée représentent 94,6% de l'inertie totale.

1/1 ✓

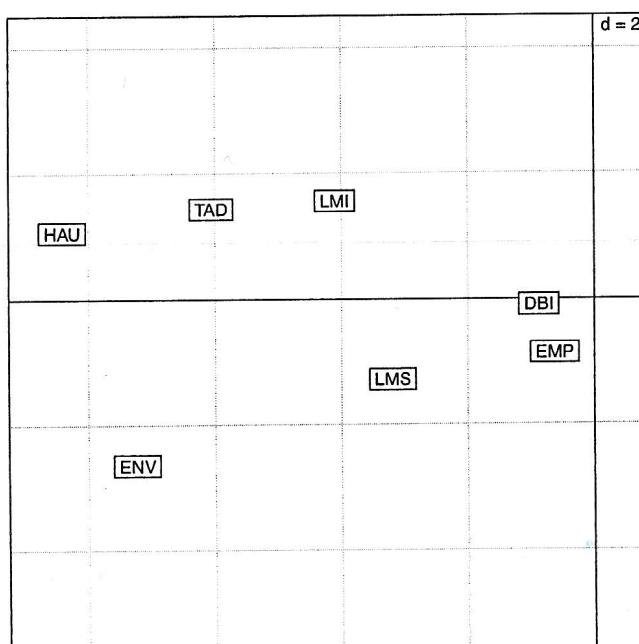
3 Morphologie et handball

Le fichier handA.txt contient la description morphologique de 28 handballeurs français de niveau international. Les mesures, toutes exprimées en centimètres, sont :

- la taille (TAD),
- la largeur interne des épaules mesurée à partir des os acromions (DBI),
- l'envergure c'est-à-dire la distance de l'extrémité d'une main à l'autre, bras en croix (ENV),
- la hauteur utilisable c'est-à-dire la hauteur mesurée bras verticaux (HAU),
- la longueur des membres inférieurs (LMI),
- la longueur des membres supérieurs (LMS),
- l'empan.

3.1 ACP centrée

- (a) On réalise une ACP centrée et on représente ci-dessous les coordonnées des variables sur le premier plan factoriel. Commenter.



Réponse : La projection sur l'axe horizontal (celui qui donne le plus d'information) fait apparaître que la hauteur utilisable, l'envergure et la taille sont très variables chez un handballer. On remarque également que toutes les variables sont à gauche du premier plan factoriel : toutes les variables sont donc plus étroitement liées. La projection sur l'axe vertical apparaît peu d'information, les variables étant trop près de l'origine pour pouvoir jager.

0/1 ✓

- (b) Donner la fonction permettant de calculer les sept variances estimées et donner les résultats avec deux chiffres après la virgule.

Réponse :

```

vares <- function(x) { sum((x - mean(x)) / length(x) - 1) }

vares(handA$TAD) 42,29   vares(handA$HAU) 76,17
"           DBI 3,50      "           LMI 22,51
"           ENV 61,83     "           LMS 12,89
                                         "           EMP 3,08

```

2 ✓
1+1

- (c) Utiliser la fonction rank sur ces variances et comparer les résultats obtenus avec le premier axe de l'analyse.

handA
 apply(handA, 2, var) → variance

Réponse :

plus la variance
est grande plus on était
loin du centre sur l'axe horizontal

L'est la variance qui
permet de placer les variables

0

3.2 ACP normée

Réaliser une ACP normée sur handA.

$$\text{acpA} \leftarrow \text{acp(handA, var = TRUE, scale = TRUE)}$$

1

- (a) Calculer le pourcentage d'inertie conservé sur le premier plan factoriel.

Réponse : $\text{handA} \leftarrow \text{read.table("handA.txt", header = TRUE)}$
 $\text{acp} \leftarrow \text{acp(handA, var = TRUE, scale = TRUE)}$

$\text{sum(acp$eig)}$
 $\text{pre} \leftarrow \text{tbon(acp$eig) / sum(acp$eig)}$
 cumsum(pre)

on conserve 84%
 de l'inertie sur le premier plan factoriel

2

- (b) Construire le cercle des corrélations et commenter la position des variables par rapport aux axes 1 et 2.

Réponse : $S.caille(acp$co, naxe = 1, gaxe = 2)$

On note d'abord que toutes les variables sont à gauche dans le cercle de corrélation et que leur projection sur l'axe horizontal sont grandes : on en déduit qu'il existe un effet taille. Si un handballeur est "bon", c'est que toutes ces caractéristiques physiques sont "bonnes". La projection sur l'axe vertical fait apparaître que les variables largeur des épaules (PBI) et empan (ETP) sont liées : les handballeurs qui ont de larges épaules ont de grandes mains.

la morphologie ne donne pas des informations sur la qualité de jeu des handballeurs

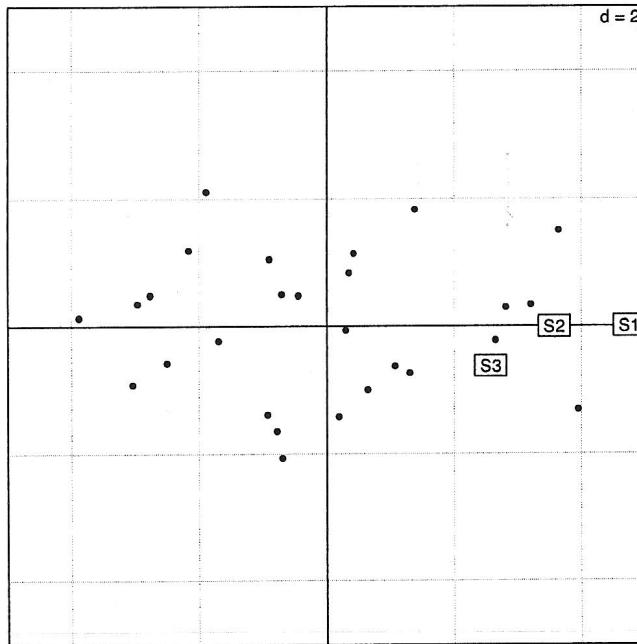
- (c) Après avoir représenter les individus sur le premier plan factoriel, commenter brièvement la position des athlètes 13, 14, 21 et 24.

Réponse : $S.Rajet(acp$li, naxe = 1, gaxe = 2)$

L'athlète 13 se caractérise par une largeur d'épaules importante et un empan important, il doit s'agir du pivot de l'équipe. L'individu 14 est le plus à gauche sur le premier plan factoriel c'est sans doute celui qui présente les meilleures caractéristiques. L'individu 21 est situé en haut à gauche : il a de bonnes aptitudes physiques mais n'est pas très large d'épaules et à un petit empan. Enfin l'individu 24, est à droite du premier plan factoriel : il ne joue donc pas au poste de pivot de l'équipe.

3.3 Individus supplémentaires

On connaît les mesures de 3 autres handballeurs appartenant à un niveau régional. On rajoute ces derniers sur la représentation graphique des individus.



- (a) Commenter leur position par rapport aux autres athlètes.

Réponse :

Les individus S1, S2 et S3 sont à droite par rapport à la plupart des autres joueurs. On peut donc dire que leur niveau régional s'oppose au niveau international des autres joueurs étudiés.

1 / 1 N+

- (b) Quelle conclusion peut-on en tirer d'un point de vue morphologique ?

Réponse :

oui
1 / 1 ✓

La conclusion qui ressort est que plus vos caractéristiques physiques sont "bonnes" plus vos chances de devenir un joueur international sont grandes.