

Analyse Discriminante

AD

Objectifs

- ▶ Étude d'un tableau Individus x Variables : Les individus sont décrits par p variables quantitatives X_1, \dots, X_p .
- ▶ Les individus sont répartis en m classes selon les modalités d'une variable qualitative Y .
- ▶ Rechercher des variables discriminantes D_i , combinaisons linéaires des X_j , non corrélées entre elles, et séparant au mieux les m classes.
- ▶ donner une représentation graphique (comme pour les méthodes factorielles telles que l'analyse en composantes principales) montrant au mieux le résultat de cette opération.

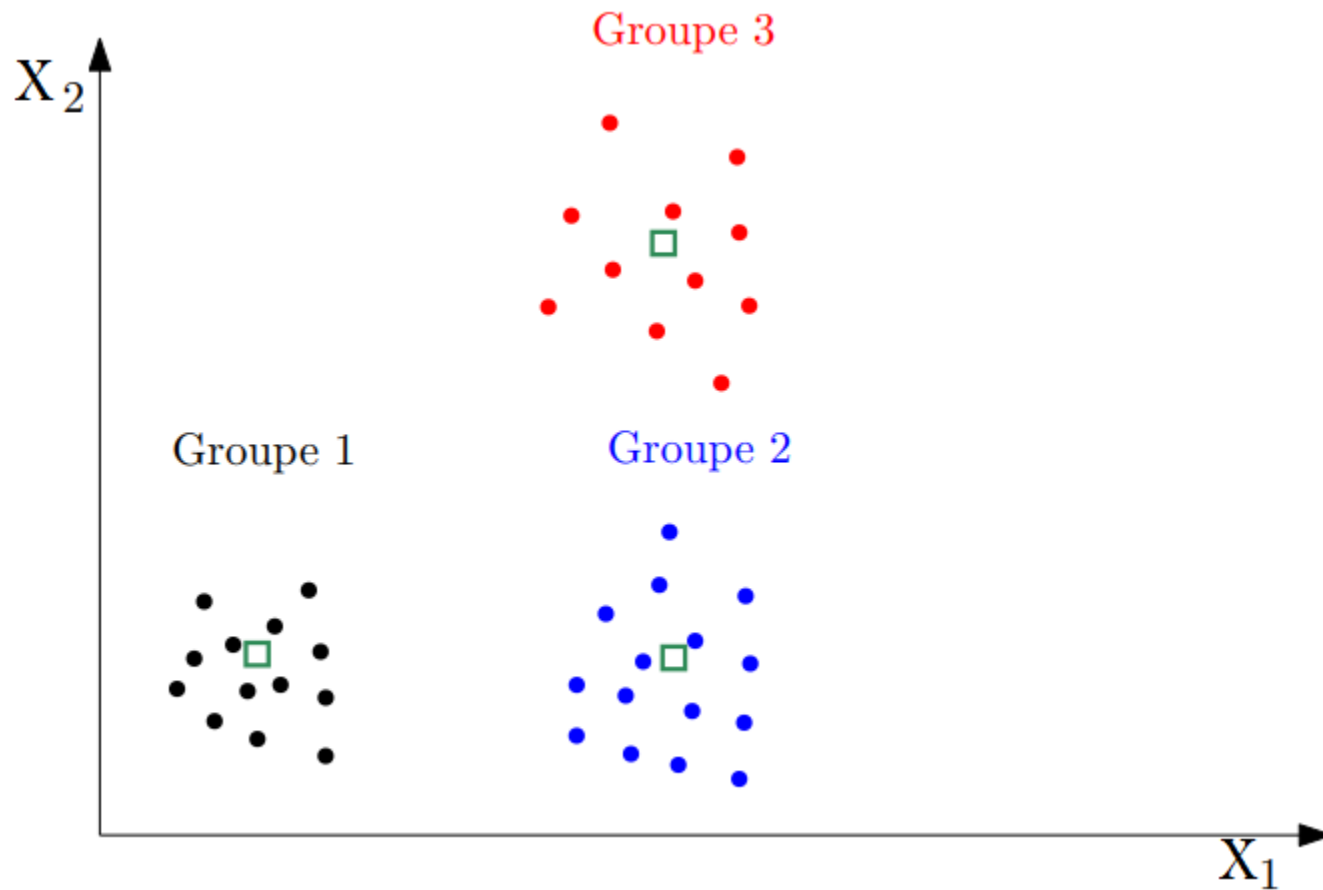
Plan du Cours

- I. Principes de l'AFD
- II. Données et définitions
- III. Premier axe discriminant
- IV. $K^{\text{ème}}$ axes discriminants
- V. Exemple d'application

Principes de l'AFD

- ✓ L'AFD est alors essentiellement descriptif.
- ✓ Les individus sont répartis en m classes selon les modalités d'une variable qualitative Y .
- ✓ Discriminer (séparer, caractériser) m groupes d'individus préalablement définis, décrits par p variables quantitatives.
- ✓ L'AFD recherche, parmi toutes les ACPs possibles sur les variables X_i , celle dont les représentations graphiques des individus discriminent "au mieux" les m classes
- ✓ Moyen : Rechercher des combinaisons linéaires des p variables initiales (axes discriminants) permettent de caractériser au mieux les groupes.
- ✓ L'AFD peut être vue comme l'ACP normée du nuage des centres de gravités des m groupes d'individus, munis du poids des groupes.

Principes de l'AFD



Les données

- Tableau X centré (sinon, on le centre)

Sur les n individus de l'échantillon, on a mesuré une variable qualitative Y à m modalités et p variables quantitatives.

$$\mathbf{X} =$$

	1 ... j ... p
1	
\vdots	\vdots
i	... x_{ij} ...
\vdots	\vdots
n	

Les données

site	YEU	BR	OP	NAG	FOI	TUB	EC	MUS	POI	LON	LART	LAR	LARM	DYEU
site1	10	65	65	107	7	76	142	1	132	214	54	47	18	11
site1	9	43	39	67	29	113	99	2	122	220	49	44	16	10
site1	6	47	71	95	11	192	121	2	129	220	49	45	17	11
site1	7	70	40	66	8	310	90	2	133	225	52	48	15	11
site1	8	59	67	100	14	289	244	1	57	168	37	37	9	9
site1	8	46	55	112	17	115	153	1	59	178	38	35	11	9
site1	7	47	36	87	16	100	162	1	59	176	40	36	11	9
site1	11	79	46	95	20	106	141	4	47	176	39	31	10	8
site2	13	80	64	155	42	192	169	3	72	182	40	39	12	10
site2	21	150	115	146	49	229	233	5	79	200	45	38	12	9
site2	12	91	84	138	22	590	220	2	80	185	43	41	12	11
site2	14	120	76	125	21	309	617	5	72	175	40	39	13	10
site2	14	142	86	135	34	523	211	10	75	189	42	39	18	10
site2	23	92	80	132	49	459	197	2	52	164	36	35	12	9
site2	13	85	64	124	20	318	191	4	86	195	41	39	16	10
site2	14	106	67	110	31	115	248	6	87	210	46	40	17	10
site3	32	224	260	314	36	107	461	3	72	181	41	36	13	9
site3	22	162	218	318	25	884	590	2	63	175	38	35	12	9
site3	31	195	208	350	73	109	809	11	49	170	39	33	12	8
site3	15	127	119	197	23	99	157	2	107	204	47	45	15	11
site3	22	160	256	282	12	102	690	3	83	190	42	44	14	9
site3	24	162	231	308	51	1031	558	2	82	194	42	39	14	10
site3	15	64	163	229	16	109	345	1	91	190	44	42	13	11

Les données

► Matrices associées

$$X = \begin{pmatrix} \tilde{X}_1 \\ \dots \\ \tilde{X}_m \end{pmatrix}$$

$$\tilde{X}_k = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \dots & \dots & \dots \\ x_{n_k 1} & \dots & x_{n_k p} \end{pmatrix}$$

P = matrice des poids
($diag = 1/n$ si poids égaux)

n_k = est le nombre d'individus qui possèdent la modalité k .

Les données

\tilde{X}_{kc} = matrice centrée correspondante

n_k = effectif du groupe k

P_k = matrice diagonale des poids des individus du groupe k
($diag = 1/n_k$ si poids égaux)

G_k = centre de gravité du groupe k = $\frac{1}{n_k} \sum_{i \in G_k} x_i$

M = matrice diagonale des poids des différents

groupes ($diag = m_k = n_k / n$ si poids égaux)

$$C = \begin{pmatrix} G_1 \\ \dots \\ G_m \end{pmatrix}$$

Définitions

- ✓ Dans l'espace des individus, le principe consiste à projeter les individus dans une direction permettant de mettre en évidence les groupes. À cette fin, Il faut privilégier la variance interclasse au détriment de la variance intraclasse considérée comme due au bruit.
- ✓ Inertie ou variance inter-classes : matrice de var-cov des p variables calculée sur le nuage des centres de gravités des m groupes. $B = C'MC = \sum_k \frac{n_k}{n} g_k g_k'$

Cas de données centrées

- ✓ Inertie ou variance intra-classes : $W = \sum m_k V_k$
$$V_k = \tilde{X}_{kc}' P_k \tilde{X}_{kc} = \frac{1}{n_k} \sum_{i \in G_k} (x_i - g_k)(x_i - g_k)'$$

Où V_k est la matrice de var-cov des p variables calculée sur les individus du groupe k

- ✓ Inertie ou variance totale : On a : $V = B + W$

Définitions

Cas particulier : Les poids sont tous égaux

$$B = \sum \frac{n_k}{n} G_k G'_k$$

$$W = \frac{1}{n} \sum n_k V_k \quad V_k = \frac{1}{n_k} \tilde{X}'_k \tilde{X}_k - G_k G'_k$$

variance totale = moyenne des variances marginales + variance des moyennes marginales.

Premier axe discriminant

❖ On cherche u_1 tel que en projection sur cet axe:

- Les centres de gravité des différents groupes soient les plus éloignés possibles (inertie INTER-classe élevée)
- Les individus d'un même groupe soient concentrés le plus possible autour de leur centre de gravité (inertie INTRA-classes faible)

Premier axe discriminant

► Objectif :

- Il s'agit de trouver une nouvelle variable, combinaison linéaire des variables explicatives, qui “discrimine” au mieux les groupes définis par les modalités de la variable à expliquer.
- Cette variable notée D est définie ici comme un vecteur de \mathbb{R}^n , combinaison linéaire des vecteurs x^1, \dots, x^p
- $D = Xu = u_1x^1 + u_2x^2 + \dots + u_px^p$
- $u = (u_1, \dots, u_p) \in \mathbb{R}^p$ est le vecteur des coefficients de cette combinaison linéaire.

► Solution : On va donc définir

- Comment mesurer que D “discrimine” bien,
- Comment trouver u pour que $D = Xu$ “discrimine” au mieux.

Premier axe discriminant

Inertie du nuage projeté :

$D_1 = X u_1$ coordonnées du nuage projeté

$$I_1 = D_1' P D_1 = u_1' X' P X u_1 = u_1' V u_1$$

$$I_1 = u_1' B u_1 + u_1' W u_1 = I_{1B} + I_{1W}$$

chercher $u_1 \in R^p /$

-Inertie inter-classes $u_1' B u_1$ maximale

-Inertie intra-classes $u_1' W u_1$ minimale

$$\Leftrightarrow \lambda_1 = \frac{u_1' B u_1}{u_1' V u_1} \quad \textbf{Maximale}$$

λ est

- la proportion de la variance de D expliquée par y
- encore appelé le rapport de corrélation entre D et y

Premier axe discriminant

u_1 est le vecteur propre unitaire de $V^{-1}B$ associé à la plus grande valeur propre λ_1 :

$$V^{-1}Bu_1 = \lambda_1 u_1$$

Définitions :

- u_1 est la direction du premier axe discriminant ou facteur discriminant. C'est le premier vecteur propre de $V^{-1}B$
- $D_1 = X u_1$ est la première variable discriminante :
 - vecteur constitué des coordonnées des n individus sur l'axe 1
- λ_1 est le pouvoir discriminant de l'axe 1

Premier axe discriminant

Propriétés :

- ▶ Un axe sera discriminant si les groupes sont bien séparés en projection.
- ▶ On aura une bonne discrimination des groupes si :
 - les centres de gravité projetés sont bien éloignés i.e. $\text{Inter}(s) = \mathbf{u}'\mathbf{W}\mathbf{u}$ est maximum,
 - les groupes projetés ne sont pas trop dispersés i.e. $\text{Intra}(s) = \mathbf{u}'\mathbf{B}\mathbf{u}$ est minimum.

Premier axe discriminant

Remarque : On constate aussi que $\mu_1 = \frac{u_1' B u_1}{u_1' W u_1}$ est maximal

La solution de ce nouveau problème est le vecteur propre unitaire de

$W^{-1}B$ associé à la valeur propre

$$\mu_1 = \frac{\lambda_1}{1 - \lambda_1}$$

Il est égal à u_1 à une constante près

- Le critère λ varie entre 0 et 1 et s'interprète facilement
- Le critère μ varie entre zéro et l'infini.

Premier axe discriminant

Prop : $\lambda_1 = \frac{u_1' B u_1}{u_1' V u_1}$ est le pouvoir discriminant de l'axe 1

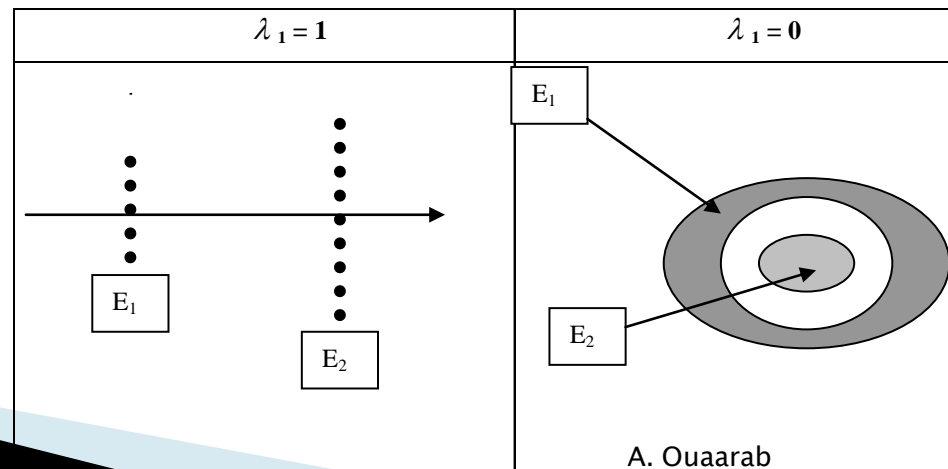
✓ $\lambda_1 \in [0, 1]$

✓ $\lambda_1 = 1$: discrimination parfaite

✓ les dispersions intragroupe sont nulles en projection sur la droite discriminante

✓ $\lambda_1 = 0$: Les centres de gravité des nuages de points sont confondus (aucune discrimination n'est possible).

✓ le meilleur axe discriminant ne permet pas de séparer les m centres de gravité



K^{eme} axes discriminants

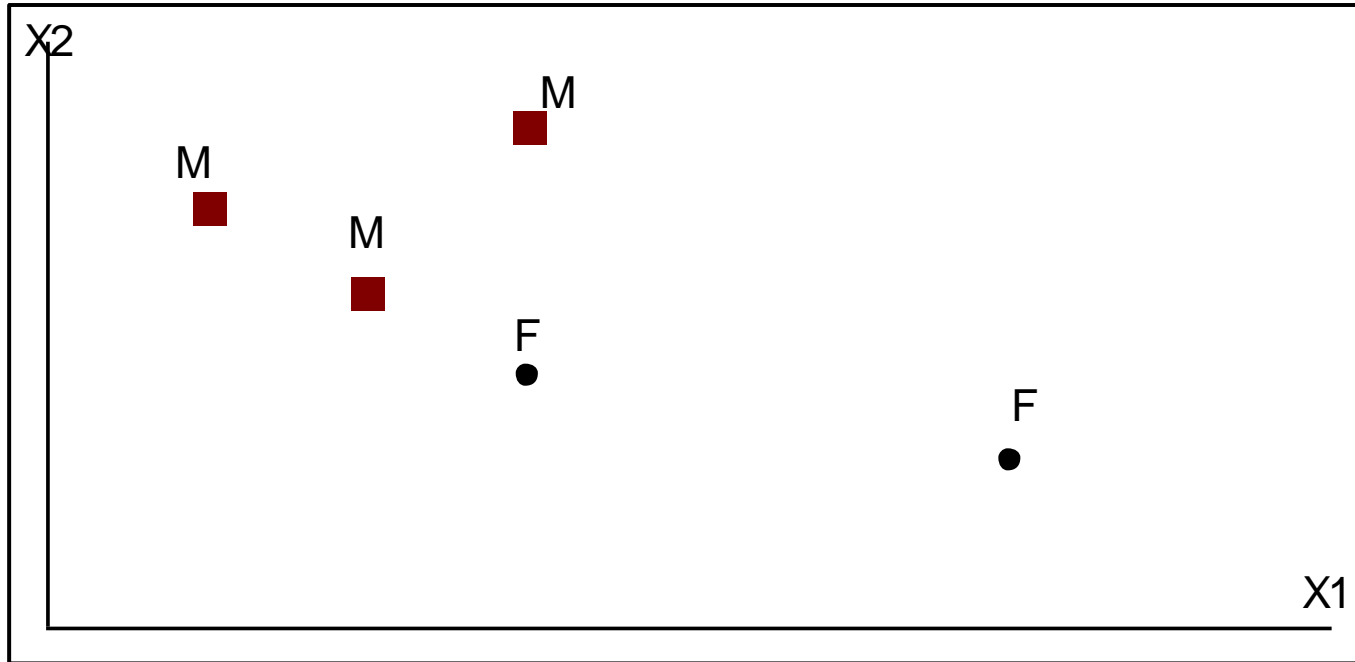
- ✓ L'AFD du tableau X s'obtient en cherchant les vecteurs propres u_k et les valeurs propres associées de $V^{-1}B$: le k^{eme} axe discriminant est le vecteur propre associé à la valeur propre de rang k de cette matrice.
- ✓ Le nombre maximum d'axes (= nombre de valeurs propres non nulles) que l'on puisse obtenir en effectuant l'AFD sur m groupes est $(m-1)$.

Exemple d'application

On observe deux variables quantitatives X_1 et X_2 sur un ensemble de $n=5$ individus de même poids, supposés répartis en deux groupes (M : masculin et F : féminin) :

Groupe	X_1	X_2
M	1	5
M	3	6
M	2	4
F	3	3
F	6	2

Exemple d'application



Exemple d'application

Grandeurs d'intérêt :

$$n_1=3, n_2=2, n=5$$

$$X = \begin{pmatrix} -2 & 1 \\ 0 & 2 \\ -1 & 0 \\ 0 & -1 \\ 3 & -2 \end{pmatrix}$$

$$M = \begin{pmatrix} 3/5 & 0 \\ 0 & 2/5 \end{pmatrix}$$

$$\tilde{X}_1 = \begin{pmatrix} -2 & 1 \\ 0 & 2 \\ -1 & 0 \end{pmatrix} \quad G_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \tilde{X}_2 = \begin{pmatrix} 0 & -1 \\ 3 & -2 \end{pmatrix} \quad G_2 = \begin{pmatrix} 1.5 \\ -1.5 \end{pmatrix}$$

Exemple d'application

Recherche de l'axe discriminant :

✓ Matrice variance totale: $V = X'X/n$ $V = \frac{1}{5} \begin{pmatrix} 14 & -8 \\ -8 & 10 \end{pmatrix}$

✓ Matrice de variance inter-classes :

$$B = C'MC \quad B = \frac{1}{5} \begin{pmatrix} 7.5 & -7.5 \\ -7.5 & 7.5 \end{pmatrix}$$

✓ Matrice de variance intra-classes :

$$V_k = \frac{1}{n_k} \tilde{X}_k' \tilde{X}_k - G_k G_k' \quad W = \frac{1}{n} \sum n_k V_k$$
$$V_1 = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad V_2 = \frac{1}{2} \begin{pmatrix} 4.5 & -1.5 \\ -1.5 & 0.5 \end{pmatrix} \quad W = \frac{1}{5} \begin{pmatrix} 6.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix}$$

Exemple d'application

Matrice à diagonaliser :

$$V^{-1}B = \frac{7.5}{76} \begin{pmatrix} 2 & -2 \\ -6 & 6 \end{pmatrix}$$

Exemple d'application

- ▶ La valeur propre non nulle de $V^{-1}B$ est = 0.79, qui est le pouvoir de discriminant de l'axe
- ▶ Le vecteur propre unitaire associé à cette valeur propre est donné par :

$$u = \begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{pmatrix}$$

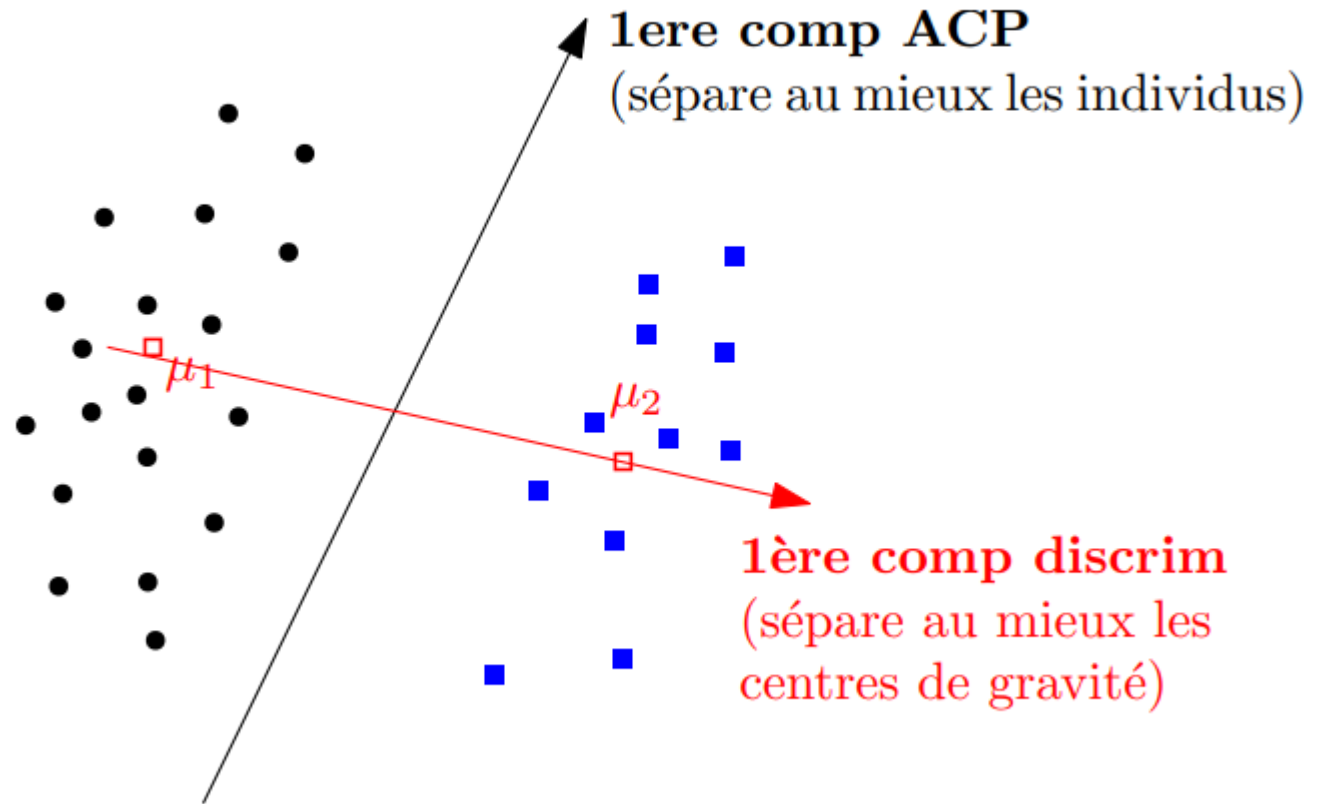
- ▶ Les coordonnées sur cet axe $D=Xu$ sont :

C =

M	5
M	6
M	1
F	-3
F	-9

AFD / ACP

Lien avec l'ACP



Exemple 2

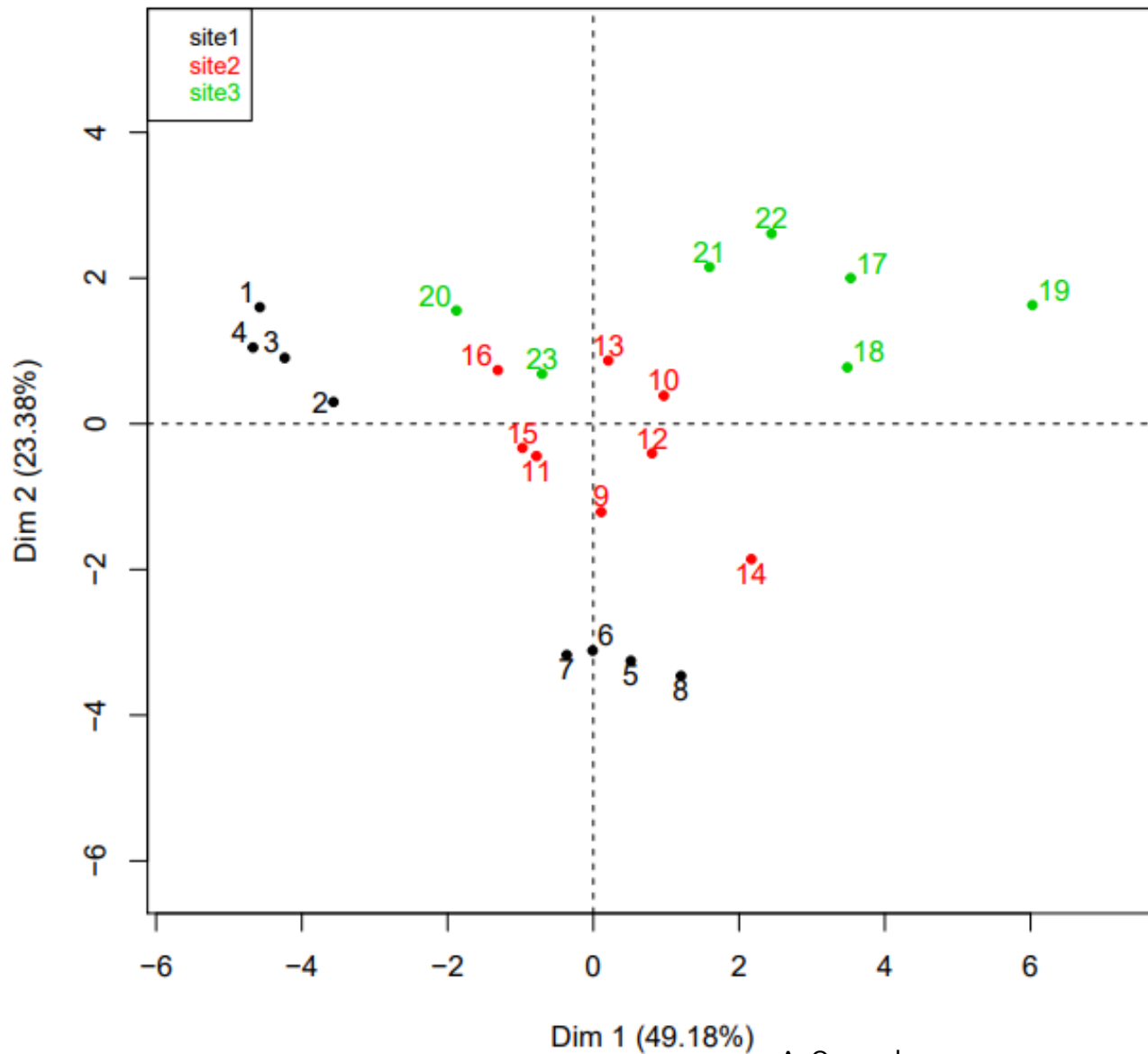
- ▶ On dispose ici de données où $n = 23$ poissons sont répartis en trois groupes selon leur site de pêche (site 1, site 2, site 3).
- ▶ La variable qualitative à expliquer Y est la variable site qui possède 3 modalités (site 1, site 2, site 3).
- ▶ Sur ces 23 poissons, on a mesuré les $p = 14$ variables quantitatives.

Objectif:

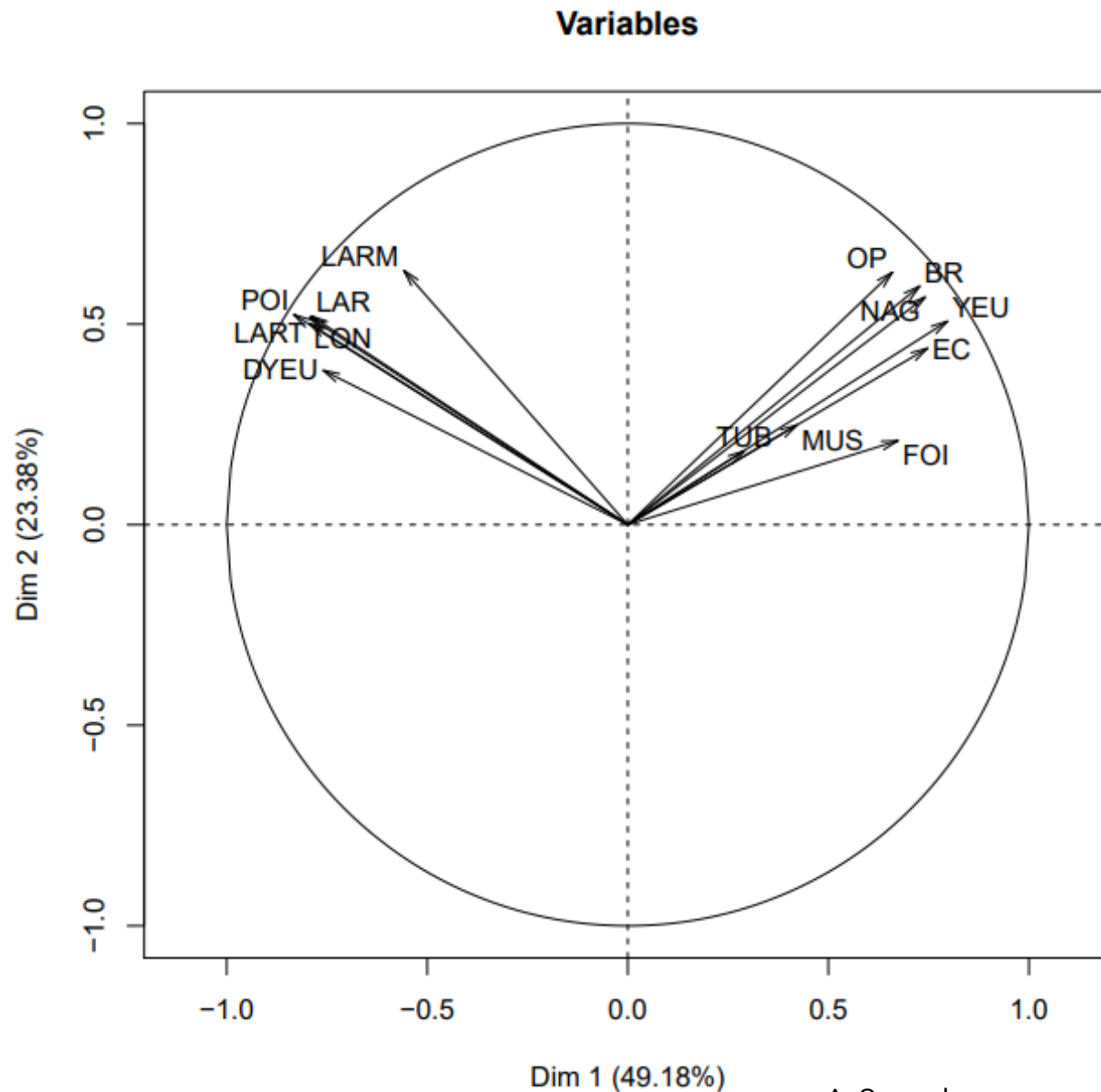
- ▶ On veut pouvoir décrire Y ou encore expliquer l'appartenance à un site de pêche, en fonction de ces 14 variables explicatives X_1, \dots, X_{14}

Premier plan factoriel de l'ACP

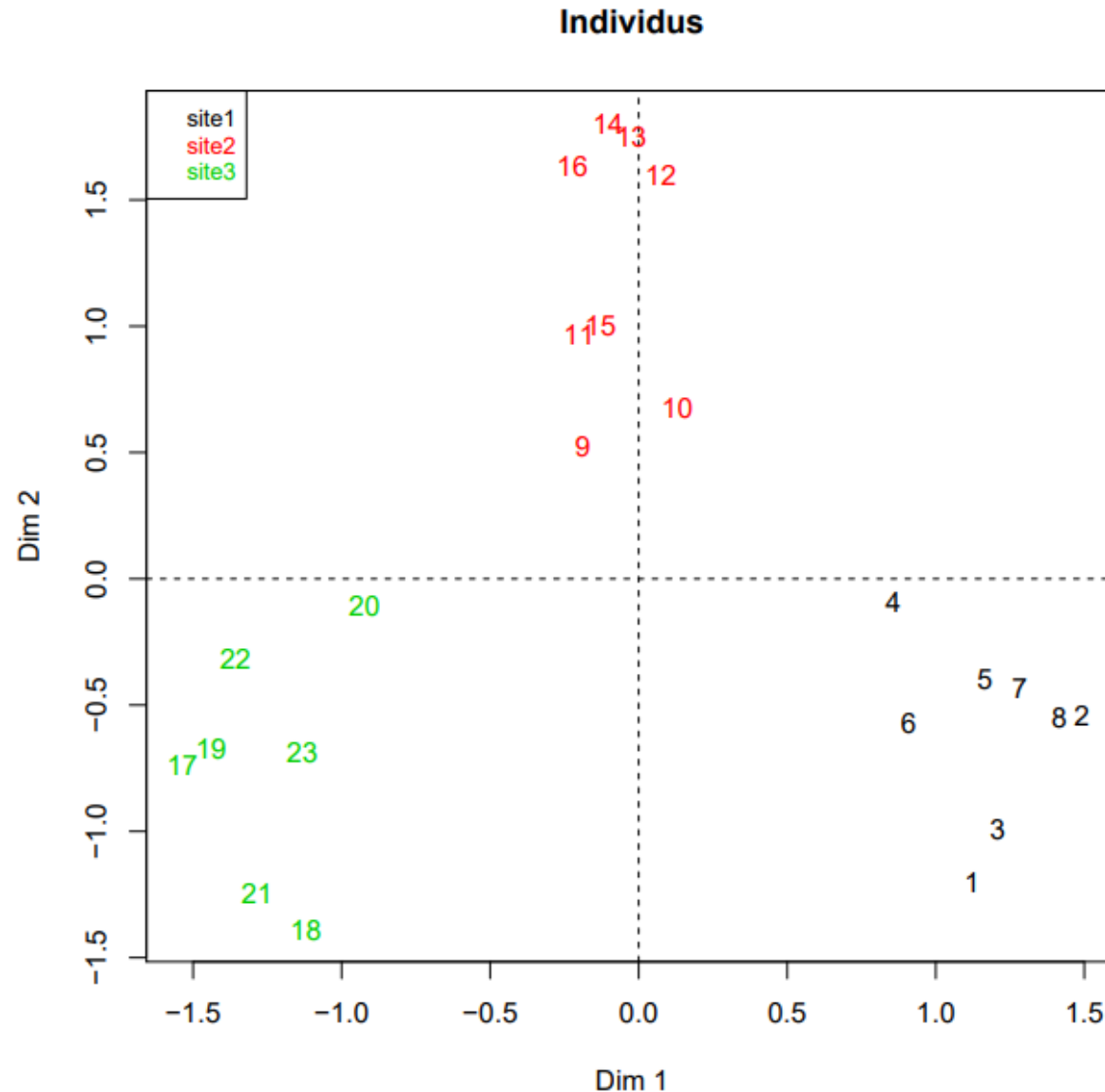
Individus



Corrélation : Axes / Variables



Premier plan factoriel de l'AFD



Corrélation : Axes/Variables

Variables

