

5. AFD – Analyse Factorielle Discriminante

Plan

1. Exemples
2. Objectifs de l'analyse discriminante
3. Introduction intuitive à l'AFD
4. Recherche des axes factoriels discriminants

5.1 – Exemples

Exemple 1 : « Hémochromatose »

- 136 patients tous atteints *d'hémochromatose* (absorption anormale de fer au niveau de l'organisme : foie, pancréas, cœur, hypophyse)
- Deux groupes : 37 atteints d'une cirrhose ; 99 indemnes
- 14 descripteurs quantitatifs ou qualitatifs

Extrait du jeu de données

74hemo																		
	SEXE	AGE	DIAG	HM	IHC	AST	ALT	GGT	FER	FRT	CS	CHFAGE	BMI	OH	CIR			
1	H	40	oui	non	1.2	1.1	1.4	54	7600	89	18.90	19.9	oui	oui		▲		
2	H	41	oui	non	0.8	0.6	4.2	42	7892	94	4.56	22.5	oui	oui				
3	H	45	oui	non	1.0	0.5	0.9	52	9000	83	17.36	24.9	non	oui				
4	F	45	oui	oui	1.5	2.2	2.4	31	7766	83	13.38	21.2	non	oui				
5	F	60	oui	oui	3.1	2.9	1.2	31	5190	95	7.02	26.4	non	oui				
6	F	54	oui	oui	1.8	1.3	1.4	30	4900	90	17.33	24.5	oui	oui				
7	H	40	non	non	5.4	4.0	5.6	31	2000	100	10.25	21.0	oui	oui				
8	H	44	oui	non	2.2	1.6	1.6	38	3882	81	14.41	18.4	oui	oui				
9	H	58	oui	non	0.9	0.6	3.6	40	5419	78	9.95	25.6	oui	oui				
10	H	38	oui	non	1.4	1.0	1.3	49	4511	97	11.50	32.4	oui	oui				
11	H	47	oui	non	1.1	0.8	2.2	28	4188	80	9.06	23.2	oui	oui				
12	H	47	oui	non	1.2	1.2	3.0	40	4000	94	9.74	23.5	oui	oui				
13	H	37	oui	non	0.6	0.7	1.3	37	4530	94	21.92	20.8	oui	oui				
14	H	48	oui	non	2.2	2.5	4.0	35	1980	81	7.71	22.4	oui	oui		▼		

Liste des variables

–	SEXE : sexe du patient
–	AGEDIAG : âge du patient au diagnostic
–	HM : Hépatomégalie
–	IHC : insuffisance hépato cellulaire
–	AST : enzymes de la cellule hépatique
–	ALT : idem
–	GGT : gamma GT
–	FER : teneur en fer dans le sang (quantité de fer par unité de volume)
–	FRT : ferritine (quantité de la protéine vecteur du fer)
–	CS : coefficient de saturation
–	CHFAGE : concentration hépatique divisée par l'âge
–	BMI : index de biomasse (poids en kg divisé par le carré de la taille)
–	OH : présence d'alcoolisme
–	CIR : présence d'une cirrhose

Objectifs

- Comprendre, expliquer, ce qui différencie les deux groupes de patients (cirrhose ou non)
- Quelles sont les variables qui permettent de caractériser/discriminer/séparer les groupes ?
- Comment, à l'aide des différents paramètres, diagnostiquer la présence d'une cirrhose ?

Exemple 2 : « Jambon »

- 21 jambons répartis en trois catégories (trois signes de qualité : LR > NF > SUP)
- 23 descripteurs sensoriels

Jambon	Qualité	IntCoulRose	HomogCoulRose	ParageGrasMuscle	TenueTranche	Humidité	Persillage	TrousFermentation	PétéchiesSang	Elasticité	Jutosité	CroquantCassant	Tendreté	Pâteux	Fibreux	Salé	Sucré	FlavViande	Assaisonnement	FalvMétal	FlavFècesUrine	FlavGras	FlavBraiséFumé	FlavLactiquePlast
P1	NF	3,9	4,5	6,2	5,8	2,9	3	2,1	0,5	4,7	3	0,4	6,5	1,4	4,8	5,5	2,4	5,6	4,7	1	1,2	4,3	1,3	0,9
P2	LR	4	5,3	6,9	5,7	2,3	3,2	0,8	0,8	4,7	2,4	0,5	5,2	1,3	4,8	5,3	1,3	6	4,5	1,3	1,4	4,6	1,3	1,1
P3	NF	4,2	4,5	5,9	8,2	2,6	3,9	0,5	2,8	3,8	3,5	0,5	5,3	1,3	5,1	5	1,5	6,4	5,5	1,1	0,6	4,7	1,9	0,8
P4	LR	4,9	5,5	5,8	7,9	1,9	4,6	0,3	0,3	5,3	2,3	1,3	5,5	1	4,6	5,9	1,2	5,5	4,2	0,9	1	4,3	2,8	0,7
⋮											⋮													⋮
P17	NF	4,7	5,8	6,8	7,4	2,2	3,2	0,6	0,5	5,8	2,4	0,3	5,7	1	4,8	5,4	1,6	5	4,5	1,8	2,1	4,5	1,5	2,5
P18	LR	5	5,3	6,3	6,5	2,8	5,2	0,6	0,7	4,6	3,3	0,4	6,6	2	4,3	5,7	1,5	5,8	4,3	1,3	1,5	5,8	0,8	1,1
P19	NF	3,9	4,2	5,2	5,5	4,9	1,3	8,5	0,5	5	5,5	2,3	6,4	2,1	3	5,5	4,6	1,9	8,2	0,6	1,2	4,3	1,2	2,2
P20	SUP	4,7	5,2	6,3	5,9	2,2	3,5	1	0,5	4,8	2,3	0,5	5,5	1	4,9	5,6	1,6	6	3,8	1,2	1,7	5	0,7	1
P21	NF	3,6	5,1	5,5	6	2,1	2,9	4,2	0,4	4,5	3,1	0,3	7,1	2	4,2	5,7	3,7	3,3	6,5	0,9	0,8	4,7	1,5	1,5

Objectifs

- Peut-on caractériser d'un point de vue sensoriel chacune des trois catégories ?
- Le profil sensoriel d'un jambon est-il lié à son signe de qualité ?

Exemple 3 : « Emmental »

- 52 emmentals répartis en deux groupes : conformes ou non
- 17 descripteurs sensoriels

Emmental	intensité du parfum	parfum propionique	parfum butyrique	texture ferme	texture souple	texture granuleuse	texture collante	texture fondante	texture caractéristique	intensité du gout	gout acide	gout salé	gout sucré	gout piquant	gout fruité	gout amer	gout caractéristique	Conformité
1	5,1	4	3,7	3,8	4,8	3,7	3,3	3	3,9	5,6	4,8	4,3	3,9	4,1	3,4	3,2	3,6	non
2	4,7	3,9	3,4	5,2	3	4	3,3	4	3,6	5,8	4,3	4,9	4	4,3	4,8	3,2	3,7	oui
3	4,7	4,2	2,9	3,7	3,3	2,8	3,4	4,7	3,9	5,7	4,3	4,7	3,9	4,2	4,9	3	4	oui
4	5,3	4,6	3,9	3	3,6	2,4	3,4	5,4	4,2	5,6	4,2	4,6	4	4,2	4,9	4,1	4	oui
5	4,3	4	2,6	3,9	4,1	3,3	2,9	4,2	3,7	5,1	4	4,3	3,9	4	4,3	3,8	4	oui
6	4,7	4	3,5	4,5	4,6	3,3	3,5	4,2	4,4	4,9	3,6	4,2	3,2	3,5	4	3,2	3,9	oui
7	3,6	3,7	2,6	4,1	4,4	3,1	3,4	4,5	4,1	4,5	3,4	3,5	3,2	3	4,2	2,6	4,3	oui
⋮								⋮										⋮
49	4,8	4,2	3,2	3,8	3,5	3,9	3,6	3,5	3,3	5,1	4,4	4,5	3,2	4,8	4,1	3,6	3,1	oui
50	4,4	3,8	3,1	3	3,8	2,5	3,8	3,9	3,2	4,8	4,2	3,9	3,5	4,5	3,4	4,7	2,4	oui
51	3,5	2,6	2,8	2,7	4,2	2,9	3,3	3,7	3,1	5	3,9	3,5	3,3	4,1	3,8	4	2,8	oui

Objectifs

- Peut-on expliquer la conformité d'un emmental par son profil sensoriel ?
- Peut-on établir une règle permettant d'affecter ou non un label de conformité sur la base de caractéristiques sensorielles ?

Exemple 4 : « Handball »

- Deux équipes de 13 joueurs chacune
- On connaît le poids et la taille de chaque joueur

Objectifs

- Peut-on expliquer ce qui différencie les deux équipes en termes de poids et de taille ?
- Les deux équipes se différencient-elles selon leur morphologie ?

hand			
	Equipe	taille	poids
1	Canada	178	82
2	Canada	180	82
3	Canada	180	86
4	Canada	180	89
5	Canada	185	79
6	Canada	185	84
7	Canada	185	93
8	Canada	186	95
9	Canada	187	90
10	Canada	188	91
11	Canada	188	95
12	Canada	188	98
13	Canada	193	101
14	Istres	189	85
15	Istres	185	84
16	Istres	186	88
17	Istres	192	90
18	Istres	193	94
19	Istres	194	102
20	Istres	185	80
21	Istres	185	85
22	Istres	188	84
23	Istres	183	78
24	Istres	187	80
25	Istres	178	73
26	Istres	188	82

Autres exemples

- **Archéologie.** Détermination du sexe d'un individu à partir de mesures morphologiques réalisées sur une partie de squelette.
- **Météorologie.** Prévission d'un risque d'avalanche à partir de mesures de températures, du niveau de neige tombée, etc.
- **Marketing.** Expliquer un comportement d'achat par diverses variables socio-démographiques.
- **Banque.** Déterminer si un client présente un risque important de non remboursement d'un prêt à partir de données telles que sa situation bancaire, sa profession, le nombre d'enfants (leur âge), propriétaire ou locataire, etc.
- **Biologie.** Construction de clefs d'identification à partir de différentes mesures : largeur des ailes, longueur des pattes, etc.

5.2 – Objectifs de l'analyse discriminante

Problématique - Objectifs

- Ces exemples entre dans le cadre statistique des méthodes de discrimination ou d'**analyse discriminante**
- En Informatique (Intelligence artificielle, apprentissage automatique), on parle de **classification supervisée**

Type de données traitées

- Une population / un échantillon de n individus répartis en K classes ou groupes connus a priori
- Chaque individu est décrit par p variables (ou paramètres, critères attributs, descripteurs)

Les deux objectifs de l'analyse discriminante

Descriptif

- Quelles sont les variables qui permettent de discriminer / différencier / séparer les différents groupes ?
- Quelles sont celles qui caractérisent ou décrivent au mieux les groupes ?

Décisionnel / Prédictif

- Construire une règle de classement (ou de décision) capable de prédire le groupe d'appartenance d'un nouvel individu, de groupe inconnu
- Détermination des performances de la règle de classement (faire le moins d'erreurs possible !)

Une grande famille de méthodes

Principales méthodes de discrimination en statistique

Mais aussi :

- Réseaux de neurones
- SVM
- Modèles de mélange
- ...

Approche géométrique	Approche probabiliste	
ANALYSE FACTORIELLE DISCRIMINANTE	MÉTHODES PARAMÉTRIQUES	
	<i>variables quantitatives</i>	<i>variables qualitatives</i>
– Cas particulier : Fonction de Fisher	– Discrimination linéaire	– Modèle multinomial
	– Discrimination quadratique	
– Méthodes de scoring (DISQUAL) ...	MÉTHODES SEMI-PARAMÉTRIQUES	
	– Discrimination logistique	
	– Classifieur bayésien naïf	
	MÉTHODES NON PARAMÉTRIQUES	
	– Techniques de lissage	
	(<i>k</i> plus proches voisins, Parzen)	
	– Segmentation (discrimination par arbre)	

Choix d'une méthode ? Il dépend

- de la nature des variables explicatives disponibles
- de l'objectif plus ou moins décisionnel / descriptif

Le tableau des données - Notations

	variables			Groupes	
	X_1	X_j	X_p	G	
individus	1			1	Groupe 1
	i	x_{ij}		k	G_k : Groupe k
	n			K	Groupe K

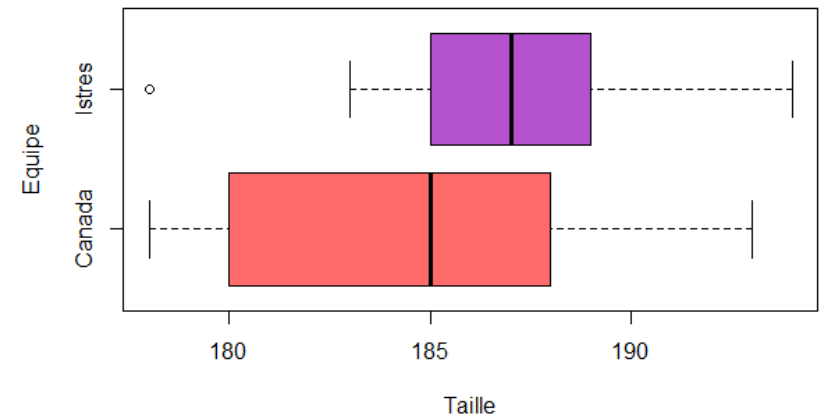
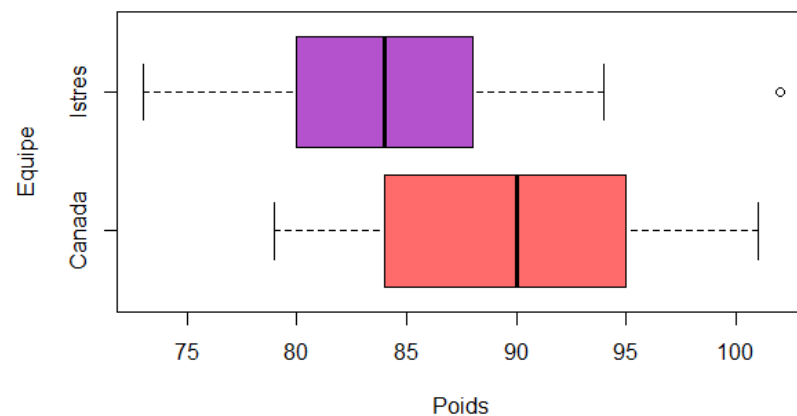
$n_k = \text{card}(G_k)$

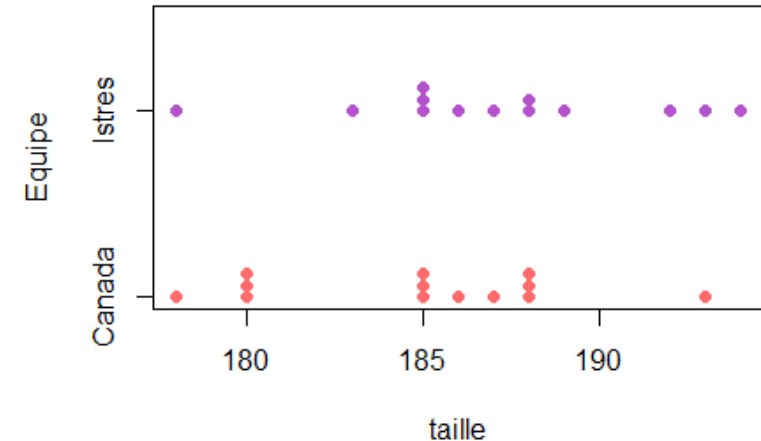
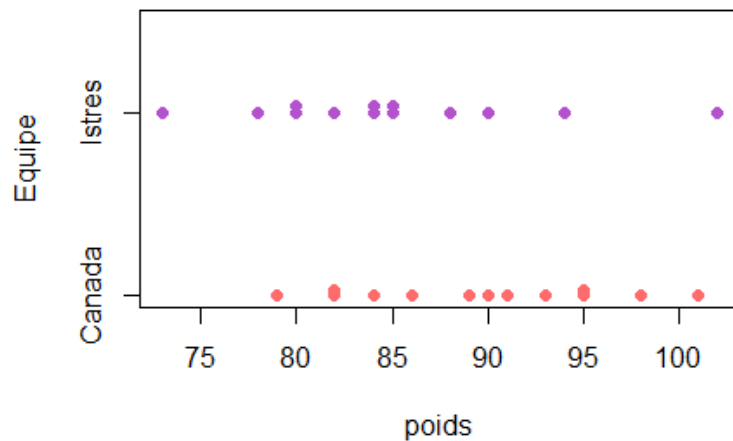
5.3 – Introduction intuitive à l'AFD

Exemple « handball »

- Comment les deux équipes se distinguent-elles du point de vue de leur morphologie ?
- La taille et/ou le poids permettent-ils de séparer (discriminer) les deux équipes ?

Étape 1. Approche univariée





Existe-t-il une différence significative entre groupes ?

Les deux équipes diffèrent-elles en moyenne selon la taille ? Selon leur poids ?

Méthodologie statistique

- Test t de différence entre deux moyennes, si deux groupes à comparer
- Analyse de la variance à un facteur si plus de deux groupes à comparer (et comparaison multiples de moyennes)

Étude de la Taille

Two Sample t-test

```
data:  taille by Equipe
t = -1.3658, df = 24, p-value = 0.1847
alternative hypothesis: true difference in
means is not equal to 0
95 percent confidence interval:
 -5.795027  1.179643
sample estimates:
mean in group Canada mean in group Istres
      184.8462              187.1538
```

Anova Table (Type II tests)

```
Response:  taille
          Sum Sq Df F value Pr(>F)
Equipe    34.62  1  1.8653 0.1847
Residuals 445.38 24
```

$$R^2 = \frac{34,62}{(34,62 + 445,38)} = 0,072$$

Étude du poids

Two Sample t-test

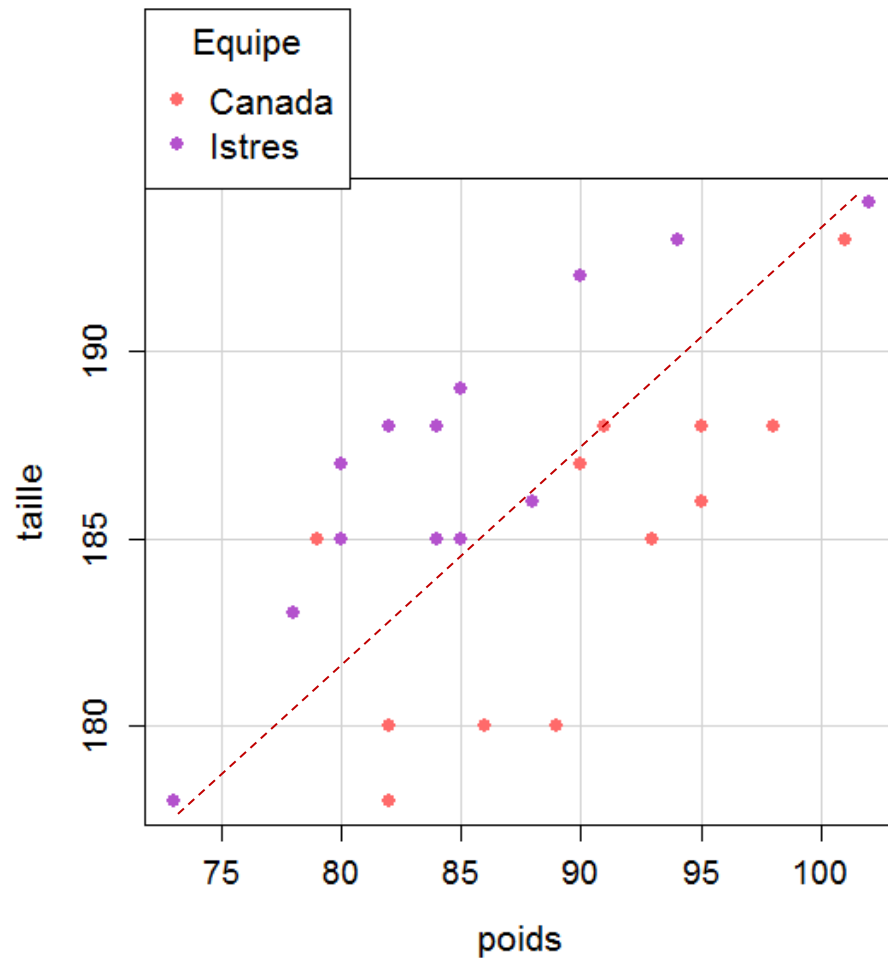
```
data:  poids by Equipe
t = 1.662, df = 24, p-value = 0.1095
alternative hypothesis: true difference in
means is not equal to 0
95 percent confidence interval:
 -1.116174 10.346943
sample estimates:
mean in group Canada mean in group Istres
      89.61538              85.00000
```

Anova Table (Type II tests)

```
Response:  poids
          Sum Sq Df F value Pr(>F)
Equipe    138.46  1  2.7621 0.1095
Residuals 1203.08 24
```

$$R^2 = \frac{138,46}{(138,46 + 1203,08)} = 0,103$$

Étape 2. Approche bivariée

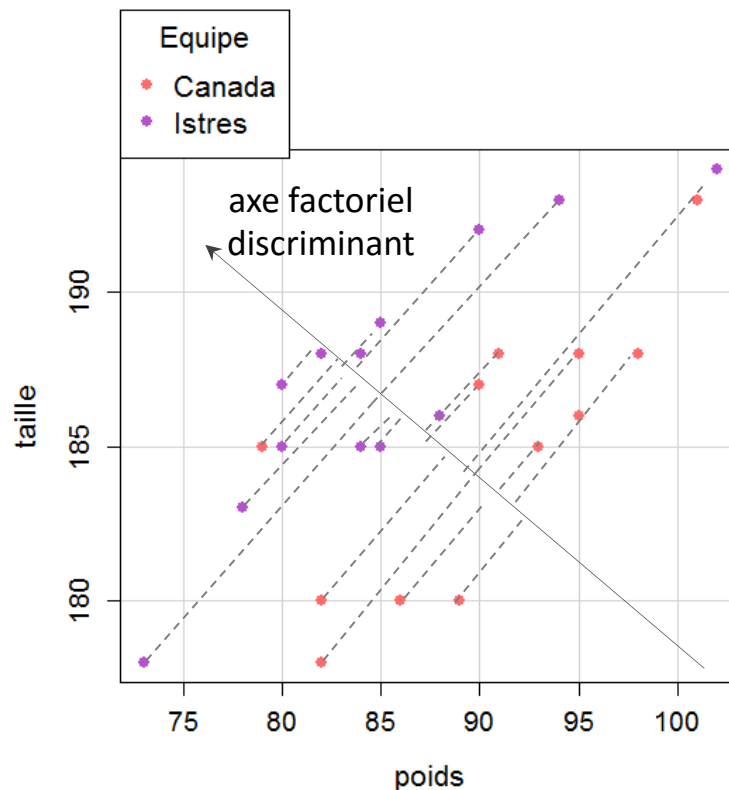


Pourtant, une séparation satisfaisante des deux groupes existe !

Interprétation...

Principe de l'analyse factorielle discriminante

- Rechercher une ou des frontières entre les groupes en se servant de l'ensemble des variables explicatives
- Construire un ou des axes factoriels discriminants, à partir de combinaisons linéaires des variables initiales



axe factoriel
discriminant

En projection sur l'axe :

Les groupes sont séparés
au mieux

Un axe discriminant : une nouvelle variable

Un axe discriminant :

- Une **variable synthétique** capable de discriminer les groupes
- Une **combinaison linéaire** des variables initiales (*Poids* et *Taille*)

Combinaison linéaire discriminante : $F_1 = 0,374.X1 - 0,232.X2 - 49,331$

Sortie du logiciel R

```
> lda(Equipe~poids+taille, data=hand)
```

```

Coefficients of linear discriminants:
              LD1
poids  -0.2324300
taille  0.3743201

```

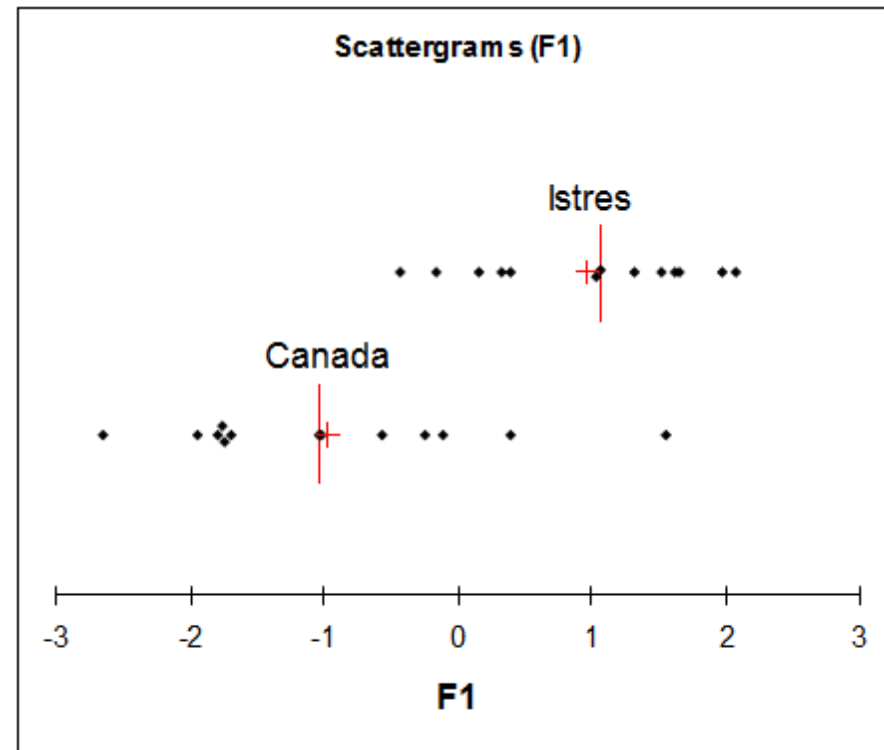
Sortie du logiciel XLSTAT

Coefficients des fonctions discriminantes canoniques :

	F1
Constante	-49,331
taille	0,374
poids	-0,232

Quel pouvoir discriminant pour F_1 ?

Equipe	taille	poids	F1
Canada	178	82	-1,761
Canada	180	82	-1,012
Canada	180	86	-1,942
Canada	180	89	-2,639
Canada	185	79	1,557
Canada	185	84	0,394
Canada	185	93	-1,697
Canada	186	95	-1,788
Canada	187	90	-0,251
Canada	188	91	-0,110
Canada	188	95	-1,039
Canada	188	98	-1,737
Canada	193	101	-0,562
Istres	189	85	1,659
Istres	185	84	0,394
Istres	186	88	-0,161
Istres	192	90	1,620
Istres	193	94	1,065
Istres	194	102	-0,420
Istres	185	80	1,324
Istres	185	85	0,162
Istres	188	84	1,517
Istres	183	78	1,040
Istres	187	80	2,073
Istres	178	73	0,331
Istres	188	82	1,982



Analyse de la variance à un facteur

- **Réponse** : F_1
- **Facteur** : variable de groupe

```
> summary(AnovaModel.2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Equipe	1	24.37	24.37	24.37	4.88e-05 ***
Residuals	24	24.00	1.00		

$$R^2 = \frac{24,37}{48,37} = 0,504$$

- La valeur de « F » obtenue est appelée « **pseudo – F** » (ne suit pas rigoureusement une loi de Fisher...)
- C'est un indicateur de l'intérêt statistique de la combinaison linéaire
- L'AFD produit des tests dits « exacts » permettant de tester l'égalité des centres de gravité des groupes (ex. : Hotelling, Wilks)

5.4 – Recherche des axes factoriels discriminants

Critère de qualité d'un axe discriminant

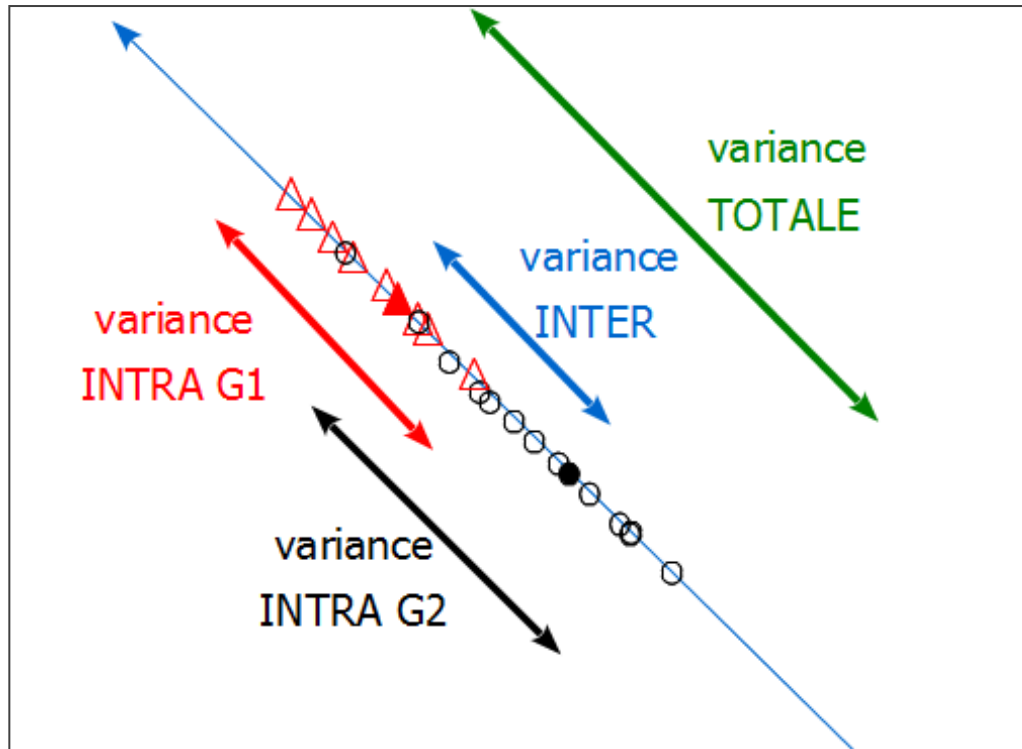
En projection

- Les groupes se **séparent au mieux**
- Les groupes sont **peu dispersés**

En terme de dispersion / variance

- la variance INTER groupes est maximale
- la variance INTRA groupes est minimale

Schématiquement



Critère optimisé

$$\max \frac{\text{Var Inter}}{\text{Var Intra}}$$

ou

$$\max \frac{\text{Var Inter}}{\text{Var Totale}} = \eta^2$$

Théorème de Huyghens

$$\text{Var Totale} = \text{Var Intra} + \text{Var Inter}$$

Variabilité Inter, Intra et Totale

Contexte

- Un ensemble de n individus répartis en K classes ou groupes
- Une variable X dont on étudié la variabilité

$$\text{Var Totale} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\text{Var Intra} = \sum_{k=1}^K \sum_{i \in G_k} (x_i - \bar{x}_k)^2$$

$$\text{Var Inter} = \sum_{k=1}^K m_k (\bar{x}_k - \bar{x})^2$$

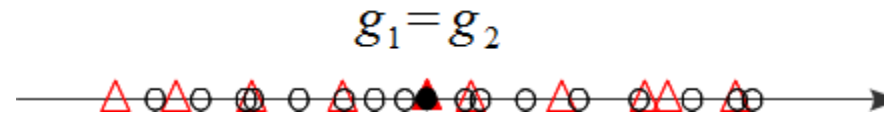
Le rapport de corrélation

$$\eta^2 = \frac{\text{Var Inter}}{\text{Var Totale}} = \frac{\text{Var Totale} - \text{Var Intra}}{\text{Var Totale}}$$

$$\eta^2 \in [0,1]$$

Les cas limites

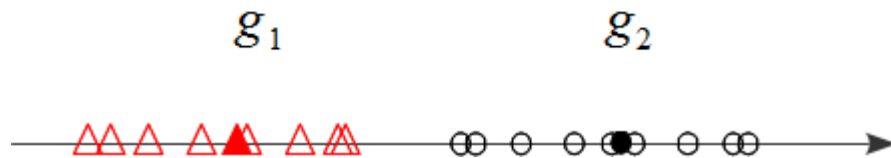
- $\eta^2 = 0$



- $\eta^2 = 1$



Une séparation parfaite, mais un $\eta^2 \neq 1$



Matrices de covariance Inter, Intra et Totale

Notations

- X = tableau de données
- x_i = description de l'individu (i)
- m_i = masse de l'individu (i)
- g = centre de gravité global
- g_k = centre de gravité du groupe (k)
- m_k = masse du groupe (k)

$$T = \sum_{i=1}^n m_i (x_i - g)(x_i - g)'$$

Matrice d'inertie **Totale**
(de variance – covariance)

$$W = \sum_{k=1}^K m_k \sum_{i \in G_k} m_i (x_i - g_k)(x_i - g_k)'$$

Matrice d'inertie **Intra**
(W = Within)

$$B = \sum_{k=1}^K m_k (g_k - g)(g_k - g)'$$

Matrice d'inertie **Inter**
(B = Between)

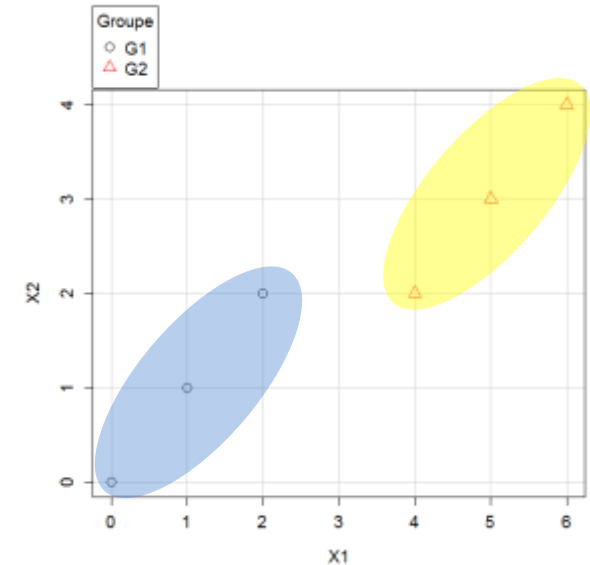
$$T = W + B$$

Illustration dans un cas simple

Contexte

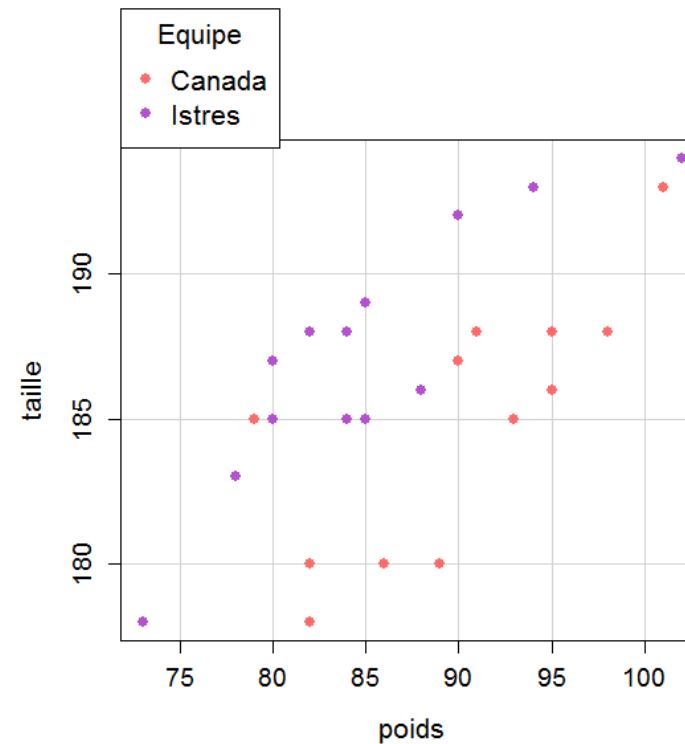
- Deux groupes G_1 et G_2
- Deux variables X_1 et X_2

X1	X2	Groupe
0	0	G1
1	1	G1
2	2	G1
4	2	G2
5	3	G2
6	4	G2



$$\begin{array}{c}
 \begin{pmatrix} V(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & V(X_2) \end{pmatrix} \\
 T
 \end{array}
 =
 \begin{array}{c}
 \begin{pmatrix} \text{Variance des } \bar{x}_{1k} & \text{Covariance } (X_1, X_2) \text{ sur cg} \\ \text{Covariance } (X_1, X_2) \text{ sur cg} & \text{Variance des } \bar{x}_{2k} \end{pmatrix} \\
 B
 \end{array}
 +
 \begin{array}{c}
 \begin{pmatrix} \text{Moyenne des } V(X_1) & \text{moyenne des cov Intra} \\ \text{Moyenne des cov Intra} & \text{Moyenne des } V(X_2) \end{pmatrix} \\
 W
 \end{array}$$

Application à l'exemple « handball »



Sortie du logiciel XLSTAT

SORTIE AFD

Matrice SSCP totale :

	taille	poids
taille	480,000	525,000
poids	525,000	1341,538

=

Matrice SSCP inter-classes :

	taille	poids
taille	34,615	-69,231
poids	-69,231	138,462

+

Matrice SSCP intra-classe totale :

	taille	poids
taille	445,385	594,231
poids	594,231	1203,077

Calcul des axes discriminants

Recherche du premier axe factoriel discriminant

- Recherche d'une direction Δu engendrée par un vecteur u
- Sur cet axe, le rapport **Inter/Intra** doit être **maximal** (ou **Inter/Total**)

Détermination des variances en projection sur Δu

- la variance TOTALE : $u'Tu$
 - la variance INTRA : $u'Wu$
 - la variance INTER : $u'Bu$
- avec $u'Tu = u'Wu + u'Bu$

Problème de maximisation. De façon équivalente,

Déterminer une direction u maximisant $\frac{u'Bu}{u'Tu}$ (1) ou $\frac{u'Bu}{u'Wu}$ (2)

Solution des problèmes de maximisation (1) et (2)

- On choisit le vecteur u associé à la plus grande valeur propre λ de $T^{-1}B$
- On choisit le vecteur v associé à la plus grande valeur propre μ de $W^{-1}B$

Relation entre les deux solutions

- Les vecteurs propres sont identiques
- Les valeurs propres sont reliées par $\lambda = \frac{\mu}{\mu+1}$ et $\mu = \frac{\lambda}{1-\lambda}$

A propos des axes discriminants

- Ils sont orthogonaux deux à deux
- Nombre d'axes discriminants = $\min(K - 1, p)$
Dans le cas le plus fréquent ($n > p > K$) : **$(K - 1)$ axes discriminants**

Illustration pour deux groupes

L'axe discriminant passe par les 2 centres de gravité
= *Fonction linéaire discriminante de Fisher*

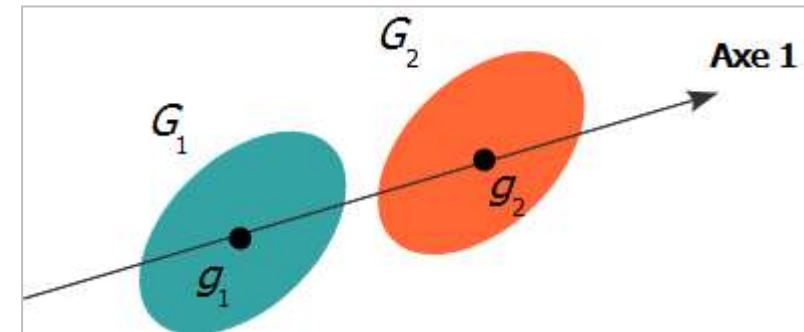
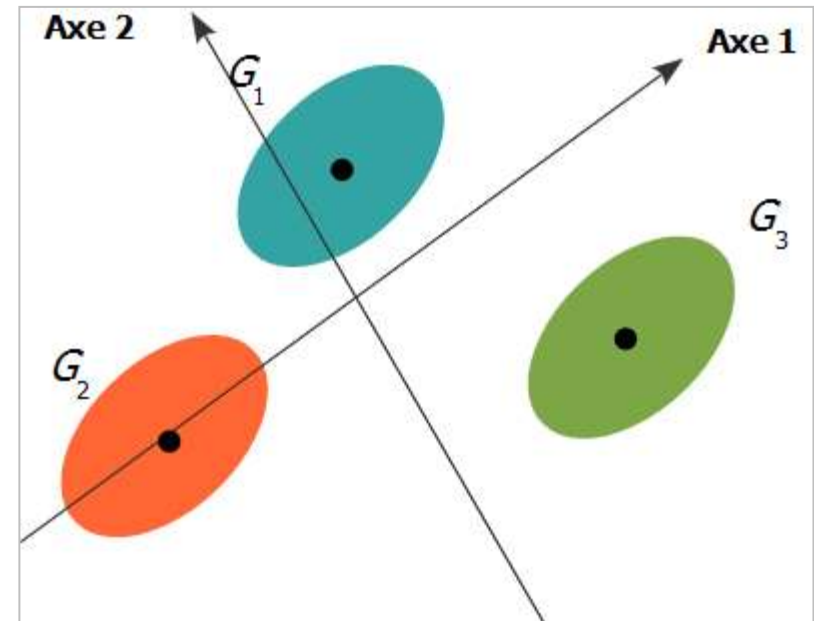


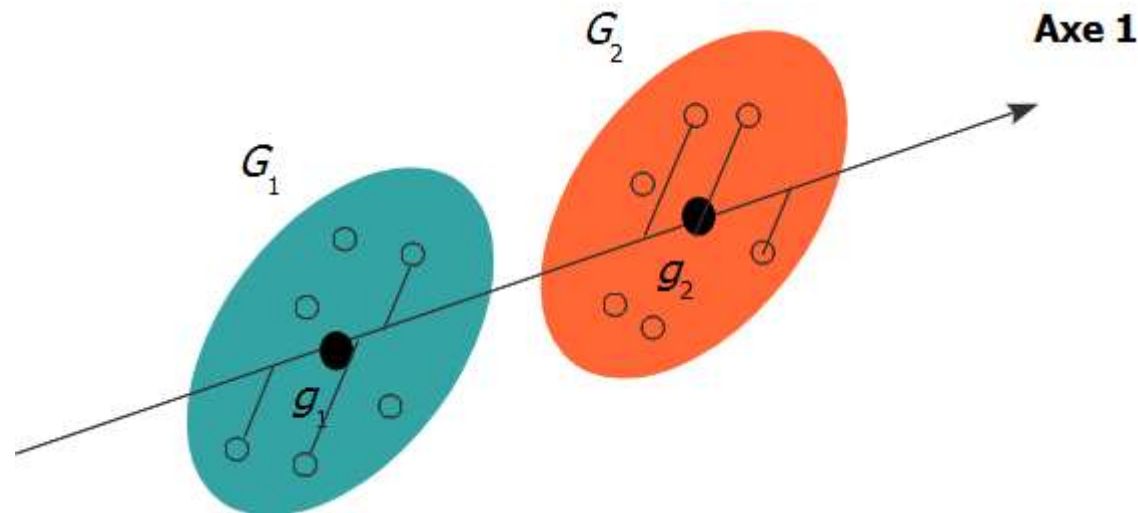
Illustration pour trois groupes

L'interprétation des axes en termes des variables initiales est réalisée au moyen du cercle des corrélations



L'AFD : une ACP particulière

- L'AFD est une ACP réalisée sur le nuage des **K centres de gravités** avec les individus en observations supplémentaires
- La métrique choisie est la matrice T^{-1} , ou W^{-1} : métrique de Mahalanobis
- L'utilisation de cette métrique revient à projeter les individus sur l'axe en accord avec la forme et l'orientation des groupes



Comment interpréter un axe discriminant ?

Interprétation des valeurs propres

- Rappel. Solution du problème de maximisation (1) : $\lambda = \frac{u'Bu}{u'Tu}$
- Propriété : $\lambda \leq 1$
- La valeur propre = mesure de séparation des groupes
= pouvoir discriminant de l'axe = rapport de corrélation

Valeurs propres :	
	F1
Valeur propre	1,016
Discrimination (%)	100,000
% cumulé	100,000

Problème de
maximisation (1)

$$\lambda = 0,504$$

Problème de
maximisation (2)

$$\mu = \frac{0,504}{1 - 0,504} = 1,016$$

L'axe en termes des variables initiales

Questions

- Quelles sont les variables initiales importantes pour la discrimination des groupes ?
- Quelles sont les variables les plus corrélées à l'axe discriminant ?

```
> F1=0.3743201*taille-0.23243*poids-49.331

> round(F1,2)
 [1] -1.76 -1.01 -1.94 -2.64  1.56  0.39 -1.70 -1.79 -0.25 -0.11 -1.04 -1.74 -0.56
[14]  1.66  0.39 -0.16  1.62  1.06 -0.42  1.32  0.16  1.52  1.04  2.07  0.33  1.98

> cor(F1,taille)
[1] 0.3783063

> cor(F1,poids)
[1] -0.4525769
```

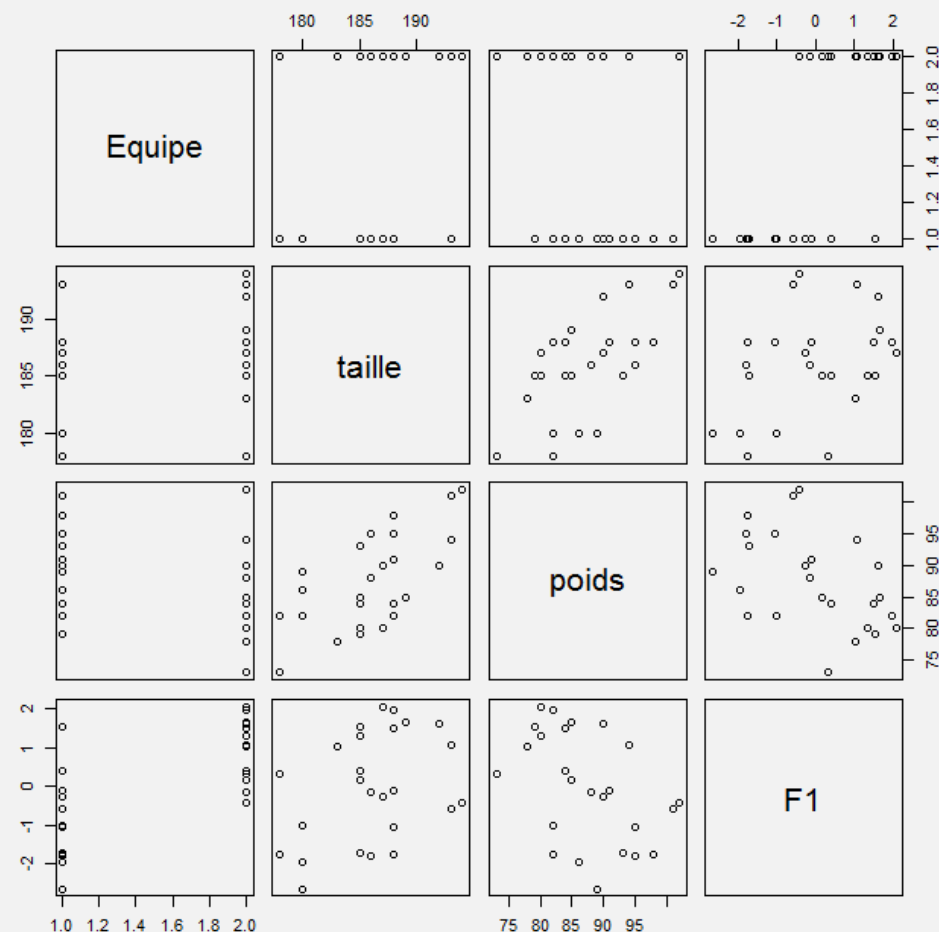
```
> rcorr.adjust(hand[,c("F1","poids","taille")], type="pearson")
```

	F1	poids	taille
F1	1.00	-0.45	0.38
poids	-0.45	1.00	0.65
taille	0.38	0.65	1.00

n= 26

P

	F1	poids	taille
F1		0.0203	0.0567
poids	0.0203		0.0003
taille	0.0567	0.0003	



5.5 – Construction d'une règle de classement

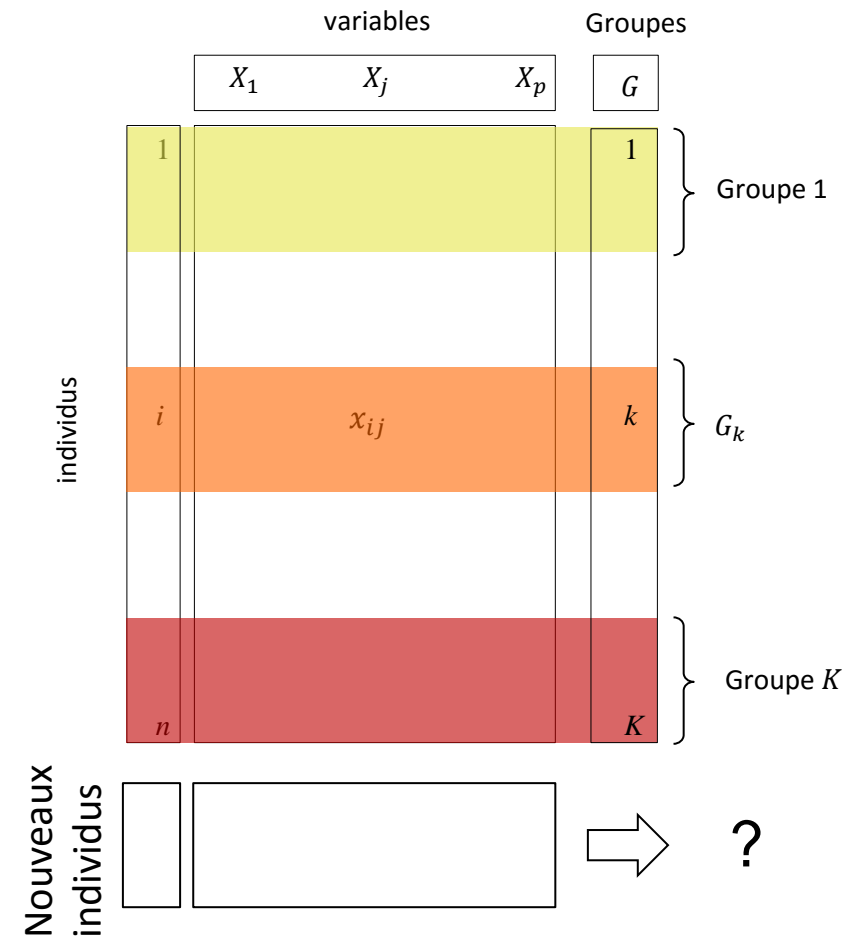
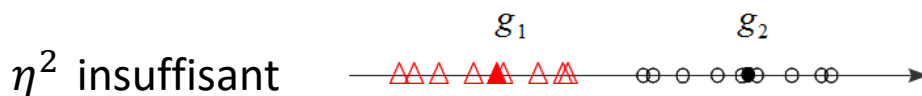
Intérêt d'une règle de classement ?

Aspect décisionnel de l'analyse discriminante

- Affecter un individu (une observation) de groupe inconnu à l'une des K classes
- Très nombreuses applications : médecine, clef d'identification en biologie, etc.

Mesure de qualité de la discrimination

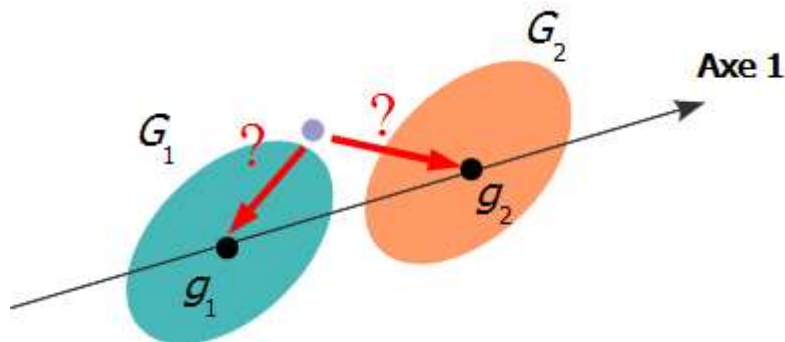
- Fournir un indicateur de qualité de discrimination en termes de taux d'individus bien classés



Construction de la règle de classement ?

Principe de la règle

On affecte un individu au groupe dont il est le **plus proche**



Quelle distance choisir ?

Calcul de la distance entre un individu (i) et le groupe G_k

$$d_{W^{-1}}^2(x_i, g_k) = (x_i - g_k)' W^{-1} (x_i - g_k)$$

W^{-1} est appelée
métrique de Mahalanobis

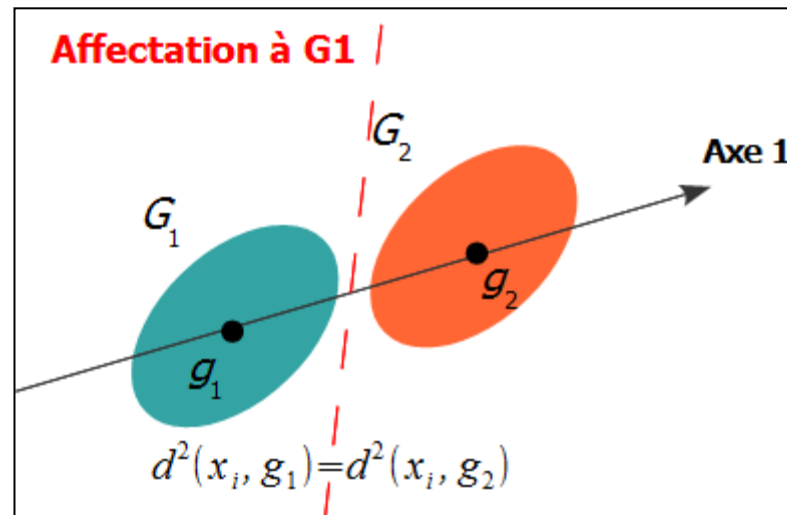
Distance entre individus (i) et le groupe G_k

Classification a priori et a posteriori, probabilités d'appartenance, coordonnées et carrés des distances :

Observation	A priori	A posteriori	Pr(Canada)	Pr(Istres)	F1	D ² (Canada)	D ² (Istres)
Obs1	Canada	Canada	0,968	0,032	-1,761	4,047	10,867
Obs2	Canada	Canada	0,877	0,123	-1,012	2,729	6,650
Obs3	Canada	Canada	0,977	0,023	-1,942	3,127	10,649
Obs4	Canada	Canada	0,994	0,006	-2,639	4,654	14,876
Obs5	Canada	Istres	0,047	0,953	1,557	8,237	2,208

Groupe d'origine Groupe d'affectation

Illustration géométrique



Les fonctions de classement

Un individu est affecté au groupe k s'il minimise

$$d_{W^{-1}}^2(x_i, g_k) = x'_i W^{-1} x_i + g'_k W^{-1} g_k - 2x'_i W^{-1} g_k$$

Fonction de classement

- Le cas de deux groupes : la fonction linéaire **discriminante de Fisher**

Un individu est affecté au groupe 1 si

$$s(x_i) = x'_i W^{-1} (g_1 - g_2) - \frac{1}{2} (g_1 - g_2)' W^{-1} (g_1 + g_2) > 0$$

Score de Fisher

- Le cas de K groupes

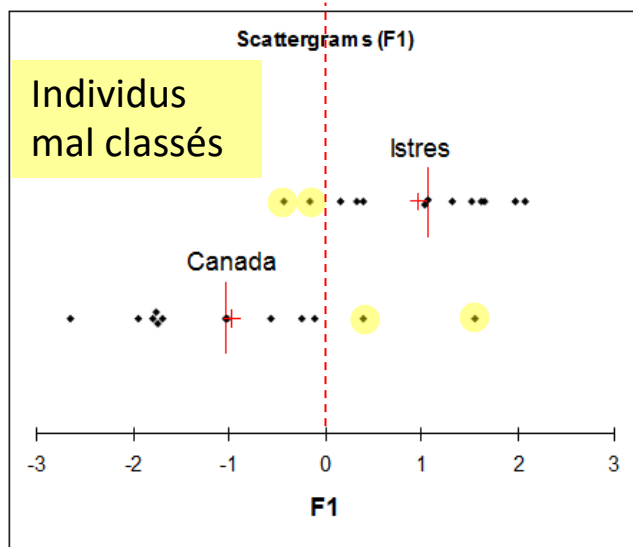
On détermine K fonctions de classement

Fonctions de classement :		
	Canada	Istres
Constante	-1642,321	-1737,854
taille	22,215	22,940
poids	-9,185	-9,635

Coordonnées des individus sur l'axe discriminant

Classification a priori et a posteriori, probabilités d'appartenance, coordonnées et carrés des distances :

Observation	A priori	A posteriori	Pr(Canada)	Pr(Istres)	F1	D ² (Canada)	D ² (Istres)
Obs1	Canada	Canada	0,968	0,032	-1,761	4,047	10,867
Obs2	Canada	Canada	0,877	0,123	-1,012	2,729	6,650
Obs3	Canada	Canada	0,977	0,023	-1,942	3,127	10,649
Obs4	Canada	Canada	0,994	0,006	-2,639	4,654	14,876
Obs5	Canada	Istres	0,047	0,953	1,557	8,237	2,208



Dans le cas de groupes d'effectifs égaux

$$d^2(x_i, g_1) < d^2(x_i, g_2) \Leftrightarrow F_1(i) < 0$$

Performance de la règle de classement ?

Principe de la règle

- La règle obtenue est-elle performante en terme de « bon classement » ?
- Est-elle capable de prédire correctement le groupe d'appartenance de nouveaux individus ?
- Problème : le groupe de nouveaux individus n'est pas connu !
- On évalue dans un premier temps la règle sur nos propres individus (de groupe connu)

Matrice de confusion

Matrice de confusion pour l'échantillon d'estimation :

de \ Vers	Canada	Istres	Total	% correct
Canada	11	2	13	84,62%
Istres	2	11	13	84,62%
Total	13	13	26	84,62%

Ces taux de bon classement sont **biaisés**
Ils sont trop **optimistes** !

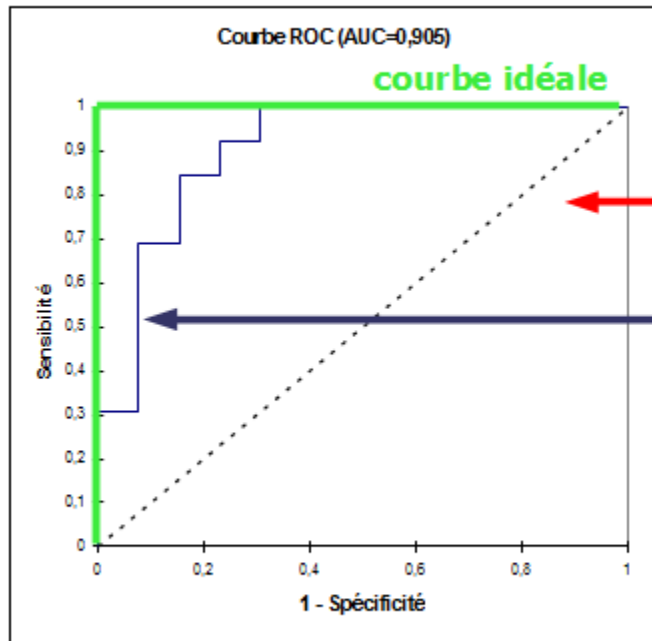
TBC = 84,62%

La courbe ROC

Receiver Operating Characteristics

Permet de visualiser la performance d'un modèle

Courbe ROC :



Ce que l'on obtiendrait par une affectation aléatoire

La courbe ROC

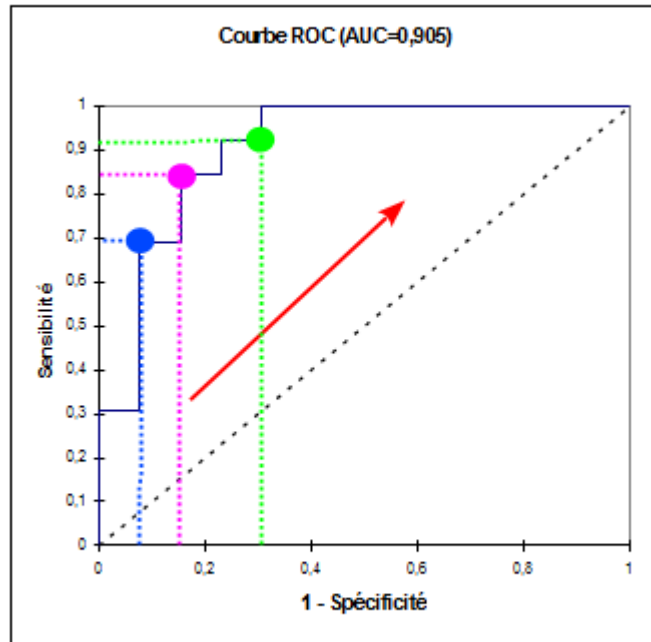
Plus elle est éloignée de la bissectrice, meilleures sont les performances du modèle

Aire sous la courbe: 0,905

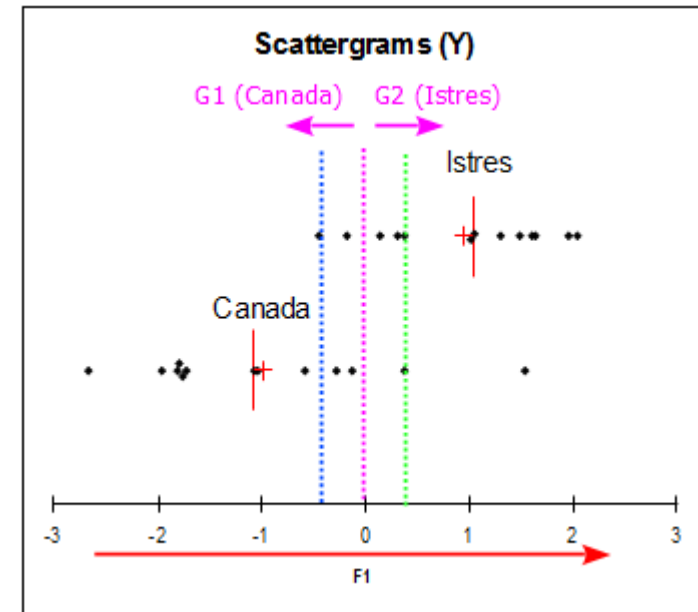
Sensibilité : proportions d'événements « positifs » (G1) bien classés

Spécificité : proportions d'événements « négatifs » (G2) bien classés

Courbe ROC :



Aire sous la courbe: 0,905



Une méthode de classement probabiliste

Principe

Connaissant la description x_i d'un individu (i), calculer directement la probabilité pour que cet individu appartienne au groupe G_k

$$P(G_k | x_i)$$

Règle de décision

On affecte l'individu (i) au groupe G_k pour lequel sa probabilité d'appartenance $P(G_k | x_i)$ est la plus élevée

Classification a priori et a posteriori, probabilités d'appartenance, coordonnées et carrés des distances :

Observation	A priori	A posteriori	Pr(Canada)	Pr(Istres)	F1	D ² (Canada)	D ² (Istres)
Obs1	Canada	Canada	0,968	0,032	-1,761	4,047	10,867
Obs2	Canada	Canada	0,877	0,123	-1,012	2,729	6,650
Obs3	Canada	Canada	0,977	0,023	-1,942	3,127	10,649
Obs4	Canada	Canada	0,994	0,006	-2,639	4,654	14,876
Obs5	Canada	Istres	0,047	0,953	1,557	8,237	2,208

Calcul des probabilités $P(G_k | x_i)$

- La règle de Bayes

Le cas de **deux groupes**

$$P(G_1 | x_i) = \frac{P(x_i | G_1) \pi_1}{P(x_i | G_1) \pi_1 + P(x_i | G_2) \pi_2}$$

π_k = probabilité *a priori* du groupe G_k

Le groupe d'affectation est donc celui pour lequel la probabilité $P(x_i | G_k) \pi_k$ est maximale

- L'hypothèse de normalité au sein des groupes

Dans le groupe G_k : x suit une loi normale $\mathcal{N}(\mu_k, \Sigma_k)$

$$f_k(x) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma_k)}} \exp\left(-\frac{1}{2}(x - \mu_k)^t \Sigma_k^{-1} (x - \mu_k)\right)$$

$$P(G_1 | x_i) = \frac{\exp(s(x_i))}{1 + \exp(s(x_i))}$$

Équivalence entre la règle probabiliste et la règle géométrique

Sous les hypothèses suivantes :

- Les descriptions des individus sont distribuées selon une loi normale multidimensionnelle dans chaque groupe : $\text{Loi}(\mathbf{x} ; G_k) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
- Égalité des matrices de variance Intra groupe $\boldsymbol{\Sigma}_k$
- Égalité des probabilités a priori π_k

La règle de décision **bayésienne** est équivalente à la règle de décision **géométrique**

On affecte au groupe 1 si et seulement si

$$d^2(x_i, g_1) < d^2(x_i, g_2) \quad \Leftrightarrow \quad P(G_1 | x_i) > P(G_2 | x_i)$$

5.5 – Validation du modèle

Existe-t-il une différence significative entre les groupes ?

Cas de deux groupes : T^2 de Hotelling et D^2 de Mahalanobis

- Objectif : tester l'égalité entre centres de gravité des deux groupes : $g_1 = g_2$
- Rappel. Le « pseudo – F » obtenu suite à l'Anova ($F_1 \sim \text{Groupe}$)
Statistique T^2 de Hotelling
- La statistique exacte de Fisher est définie par :

$$\frac{(n_1 + n_2 - 2) - (p - 1)}{p} \times \frac{T^2}{(n_1 + n_2 - 2)} \sim F(p; n_1 + n_2 - p - 1)$$

- Liaison entre T^2 et D^2

$$D^2 = (g_1 - g_2)^t W^{-1} (g_1 - g_2) \qquad T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2$$

Cas de K groupes : la statistique du Λ de Wilks

- Objectif : tester l'égalité entre les K centres de gravité : $g_1 = g_2 = \dots = g_K$
- La statistique Λ de Wilks est définie par

$$\Lambda = \frac{|W|}{|V|}$$

- Conditions d'application :
 - Normalité des données
 - Matrices de covariance des groupes identiques (Test de Box)
- Dans le cas de deux groupes, les tests du T^2 , du D^2 et du Λ sont équivalents

Test du Lambda de Wilks (approximation de Rao) :

Lambda	0,496
F (Valeur observée)	11,681
F (Valeur critique)	3,422
DDL1	2
DDL2	23
p-value	0,00032
alpha	0,05

$$F = \frac{(13+13-2)-(2-1)}{2} \times \frac{24,377}{(13+13-2)} = 11,68$$

La sélection des variables explicatives

Problématique

- Lorsque les variables explicatives sont nombreuses, risque de produire un modèle « sur paramétré »
- La règle de décision sera efficace sur les données ayant produit la règle mais très peu pour de nouvelles données !
- Il est facile de produire une règle discriminante efficace à partir d'un modèle comportant beaucoup de paramètres, et peu d'observations...

Solution

- Principe : introduire, pas à pas, les variables dans la règle
- À chaque étape : la variable à introduire présente-t-elle un intérêt ?
- Test de variation du Lambda de Wilks (\sim loi de Fisher) : « gagne-t-on en pouvoir de séparation des groupes avec une nouvelle variable ? »
- Option stepwise : à chaque étape, l'intérêt d'une variable est remis en question

Comment estimer un taux d'erreurs non biaisé ?

Problématique

- Les performances « par défaut » de la règle sont optimistes !
- La règle est évaluée à partir des données même qui ont conduit à son élaboration (méthode dite de resubstitution)
- Il faudrait pouvoir l'évaluer sur de nouveaux individus

Les solutions

- Méthode de échantillon test
- Validation croisée (LOO)
- Technique du bootstrap

- **Méthode de échantillon test**

On partage l'ensemble des observations en deux parties

- Un échantillon d'apprentissage (2/3 à 3/4 des observations) pour la construction de la règle de décision
- Un échantillon test (1/4 à 1/3 des observations) pour estimer le taux d'erreur de façon non biaisée

- **Validation croisée (LOO)**

On construit la règle de décision sur un échantillon de taille $(n - 1)$

L'observation restante est classée à l'aide de la règle obtenue

Le taux d'erreur final est la moyenne des taux obtenus sur les n échantillons possibles de taille $(n - 1)$

Équivalent à n discriminations avec un échantillon test d'une unité

- **Technique du bootstrap**

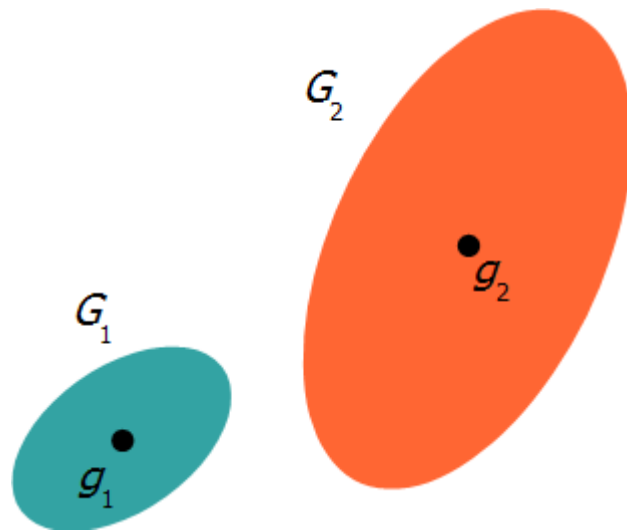
On estime le taux d'erreur en réalisant la moyenne de taux d'erreur obtenus sur un grand nombre d'échantillons bootstrap

échantillon bootstrap = n observations tirées avec remise

AFD linéaire ou quadratique

Problématique

Les groupes ne présentent pas des matrices de covariance INTRA similaires



Test de l'égalité des matrices de covariance
Test de Box ou Kullback

Conséquence : la règle discriminante obtenue est de nature **quadratique**