

UNIWIN VERSION 9.7.0

ANALYSE DISCRIMINANTE PAS A PAS

Révision : 02/09/2023

Définition.....	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Fichier IRIS	3
L'option Rapports	4
L'option Graphiques	5
Exemple 2 : Fichier INFARCT	5

Définition

L'Analyse Discriminante Pas à Pas (DISP) permet de sélectionner à partir d'un ensemble de variables quantitatives et d'une variable qualitative découpant la population en plusieurs groupes (2 ou plus), le sous-ensemble des variables quantitatives les plus explicatives des groupes qui pourront alors être utilisées pour définir des fonctions discriminantes robustes.

La méthode proposée permet de préciser les valeurs des Fishers pour l'entrée et la sortie d'une variable, les critères d'arrêt de l'algorithme (variation minimale du Lambda de Wilks si plus de 2 groupes ou de la distance D2 de Mahalanobis si 2 groupes, nombre maximal de variables à sélectionner, nombre maximal d'itérations).

La méthode itérative peut être "manuelle" (affiche toutes les étapes) ou "automatique" (affiche uniquement le tableau final). L'algorithme est formé d'une succession d'étapes ascendantes (entrée d'une variable) et descendantes (sortie d'une variable).

Un rapport général de synthèse est proposé : il permet de visualiser toutes les étapes de l'algorithme.

Cette méthode peut être complétée par une Analyse Discriminante Bayésienne (ADB) ou par une Analyse Factorielle Discriminante (AFD).

Entrée des données

Cliquons sur l'icône DISP dans le ruban Expliquer. La boîte de dialogue montrée ci-dessous s'affiche :

Analyse discriminante pas à pas

Facteur de classement :

Variables quantitatives :

(Libellés des variables quantitatives :)

Critères de sélection

Fisher d'entrée
2,072

Fisher de sortie
1,070

Affichage

Modèle final

Toutes les étapes

Critères d'arrêt

Variation minimale D2 ou lambda	Nombre maximum de variables	Nombre maximum d'itérations
5	100	50
1 %	1	10
2 %	2	20
3 %	3	30
4 %	4	40
5 %	5	50
10 %	6	60

Ok Annuler Sélection Supprimer Aide

Cette boîte de dialogue permet de définir le facteur de classement (variable qualitative), les variables quantitatives, les libellés optionnels des variables quantitatives, les critères de sélection des variables (Fisher) en entrée et en sortie, ainsi que les critères d'arrêt de l'algorithme itératif (pourcentage minimal de variation de la distance D2 de Mahalanobis ou du Lambda de Wilks, nombre maximum de variables, nombre maximum d'itérations).

Le Fisher d'entrée doit être supérieur au Fisher de sortie. Les valeurs proposées par défaut correspondent aux seuils à 85% (2,072) et à 70% (1,07) de la loi de Fisher $F(1, \text{infini})$.

Données manquantes

Les données manquantes ne sont pas autorisées pour le facteur de classement. Elles sont autorisées pour les variables quantitatives.

Exemple 1 : Fichier IRIS

Nous utiliserons le fichier IRIS pour illustrer cette procédure. Ce fichier contient pour 150 iris de trois espèces différentes les mesures des quatre caractéristiques suivantes exprimées en millimètres : longueur du sépale, largeur du sépale, longueur du pétale, largeur du pétale. Les trois espèces sont : Iris Setosa, Iris Versicolor et Iris Virginica

Cliquons sur l'icône DISP dans le ruban Expliquer.

numiris
lonsepal
larsepal
lonpetal
larpetal
mesures
nomesp
codesp1
codesp2

Facteur de classement :
codesp2

Variables quantitatives :
lonsepal
larsepal
lonpetal
larpetal

(Libellés des variables quantitatives :)
mesures

Critères de sélection
Fisher d'entrée
2,072
Fisher de sortie
1,070

Critères d'arrêt

Variation minimale D2 ou lambda	Nombre maximum de variables	Nombre maximum d'itérations
5	100	50
1 %	1	10
2 %	2	20
3 %	3	30
4 %	4	40
5 %	5	50
10 %	6	60

Affichage
 Modèle final
 Toutes les étapes

Ok Annuler Sélection Supprimer Aide

Choisissons les variables *lonsepal* à *larpetal* comme variables quantitatives, la variable *codesp2* comme facteur de classement et la variable *mesures* comme variable contenant les libellés des variables quantitatives.

Choisissons l'affichage de toutes les étapes et laissons les autres champs aux valeurs par défaut.

Cliquons sur le bouton Ok. Après quelques instants, la fenêtre suivante 'affiche :

	1	2	3	4	5	6	7	8
1								
2	(C) UNWIN version 9.3.5							
3								
4	DATE : 13/03/2022							
5	ORDINATEUR : LAPTOP-8CVD8RT1							
6	UTILISATEUR : Christian							
7	FICHIER(S) DE DONNEES OUVERT(S) : IRIS.SGD							
8								
9	RESULTATS DE L'ANALYSE DISCRIMINANTE PAS A PAS							
10								
11	Sélection :							
12	Aucune							
13								
14	Nombre d'observations complètes : 150							
15								
16	Variables quantitatives :							
17	lonsepal							
18	larsepal							
19	lonpetal							
20	larpetal							
21								

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML.

L'impression des rapports fait appel à la procédure 'Aperçu avant impression'. Pour des informations sur cette procédure, voir le 'Manuel de l'Utilisateur'.

Voici trois exemples du rapport pour notre analyse : Explorateur, Général, HTML.

	1	2	3	4	5	6	7	8
1								
2	Itération 1 - Ascendante - Variables							
3								
4		Variable	Fisher	Prob > F				
5	lonsepal	119,26450	0,00000					
6	larsepal	49,16004	0,00000					
7	lonpetal	1180,16118	0,00000					
8	larpetal	960,00715	0,00000					
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								

	1	2	3	4	5	6	7	8	9	10	11	12	13
160	lonpetal	35,59017	0,00000										
161	larpetal	24,90433	0,00000										
162													
163	Itération 4 - Descendante - Elimination												
164													
165													
166													
167	Variable éliminée :												
168													
169	Aucune.												
170													
171													
172													
173	Tableau de synthèse												
174													
175		Itération-Variable	Fisher	Prob > F	Lambda	Pct Lambda							
176	1	lonpetal	1180,16118	0,00000	0,05863	0,00000							
177	2	larsepai	43,03545	0,00000	0,03688	37,08819							
178	3	larpetal	34,56869	0,00000	0,02496	32,28646							
179	4	lonsepai	4,72115	0,01033	0,02344	6,15365							
180													

(C) UNIWIN version 9.3.5

DATE : 13/03/2022
 ORDINATEUR : LAPTOP-8CVD8RT1
 UTILISATEUR : Christian
 FICHIER(S) DE DONNEES OUVERT(S) : IRIS.SGD

RESULTATS DE L'ANALYSE DISCRIMINANTE PAS A PAS

Sélection :
 Aucune

Nombre d'observations complètes : 150

Variables quantitatives :
 lonsepai
 larsepai
 lonpetal
 larpetal

Facteur de classement : codesp2
 Nombre de classes : 3

Fisher d'entrée : 2,072
 Fisher de sortie : 1,07

Nombre maximum de variables : 4
 Nombre maximum d'itérations : 50

L'option Graphiques

Il n'y a pas de graphiques pour cette procédure.

Exemple 2 : Fichier INFARCT

Pour ce deuxième exemple, nous utiliserons le fichier INFARCT. Ce fichier contient des informations concernant 81 victimes d'un infarctus du myocarde.

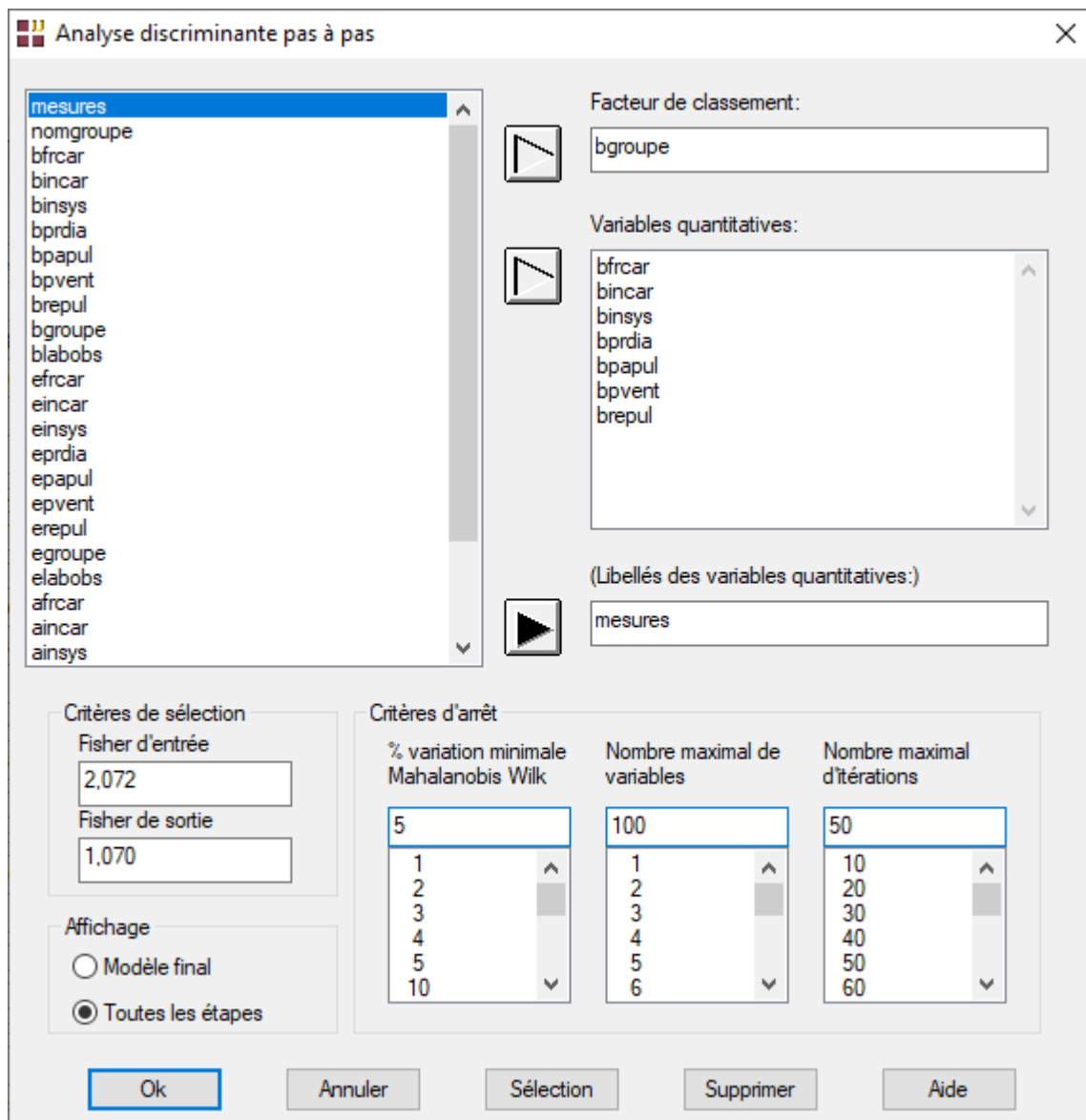
Les variables mesurées sont :

- *bfrcar* fréquence cardiaque
- *bincar* index cardiaque
- *binsys* index systolique
- *bprdia* pression diastolique
- *bpapul* pression artérielle pulmonaire
- *bpvent* pression ventriculaire
- *brepul* résistance pulmonaire

La variable *mesures* contient les libellés des sept variables.

La variable qualitative *bgroupe* indique par ses deux codes les personnes décédées (codées 1) ou vivantes (codées 2).

Cliquons sur l'icône DISP dans le ruban Expliquer. La boîte de dialogue montrée ci-dessous s'affiche. Choisissons les variables et demandons l'affichage de toutes les étapes.



Après avoir cliqué sur OK, l'écran 'Rapports et Graphiques s'affiche.

Dans ce deuxième exemple, seules deux mesures sont sélectionnées : *incar* et *papul*, les autres variables ayant des valeurs du Fisher inférieures au seuil d'entrée choisi.

Quelques écrans des résultats obtenus sont montrés ci-après.

Rapports et Graphiques

Rapport DISP

- Ascendante - Variables
- 1 - Ascendante - Sélection
- 2 - Ascendante - Variables
- 2 - Ascendante - Sélection
- 2 - Descendante - Variables
- 2 - Descendante - Elimination
- 3 - Ascendante - Variables
- 3 - Ascendante - Sélection
- Tableau de synthèse

	1	2	3	4	5	6	7	8
1								
2	Itération 1 - Ascendante - Variables							
3								
4		Variable	Fisher	Prob > F				
5		frcar	5,62238	0,02017				
6		incar	69,96208	0,00000				
7		insys	64,03969	0,00000				
8		prdia	33,64305	0,00000				
9		papul	27,55587	0,00000				
10		pvent	2,14870	0,14668				
11		repu1	63,59142	0,00000				
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								

Rapport Explorateur /

Rapports et Graphiques

Rapport DISP

- 1 - Ascendante - Variables
- 1 - Ascendante - Sélection
- 2 - Ascendante - Variables
- 2 - Ascendante - Sélection
- 2 - Descendante - Variables
- 2 - Descendante - Elimination
- 3 - Ascendante - Variables
- 3 - Ascendante - Sélection
- Tableau de synthèse

	1	2	3	4	5	6	7	8
1								
2	Tableau de synthèse							
3								
4		Itération-Variable	Fisher	Prob > F	D2	Pct D2		
5	1	incar	69,96208	0,00000	3,45544	0,00000		
6	2	papul	15,12324	0,00021	4,88193	41,26224		
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								

Rapport Explorateur /

Rapports et Graphiques

(C) UNIWIN version 9.3.5

DATE : 13/03/2022
 ORDINATEUR : LAPTOP-8CV08RT1
 UTILISATEUR : Christian
 FICHIER(S) DE DONNEES OUVERT(S) : INFARCT.SGD

RESULTATS DE L'ANALYSE DISCRIMINANTE PAS A PAS

Sélection :
 Aucune

Nombre d'observations complètes : 81

Variables quantitatives :
 frcar
 incar
 insys
 prdia
 papul
 pvent
 repul

Facteur de classement : bgroupe
 Nombre de classes : 2

Fisher d'entrée : 2,072
 Fisher de sortie : 1,07