

UNIWIN VERSION 9.7.0

ANALYSE DISCRIMINANTE BAYESIENNE

Révision : 02/09/2023

Définition	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Fichier IRIS3	3
L'option Rapports	8
L'option Graphiques	9
Exemple 2 : Fichier INFARCT2	13
L'option Rapports	16
L'option Graphiques	18
Exemple 3 : Fichier BORDEAUX	20
Exemple 4 : Fichier TITANIC	23
Les variables internes créées par la procédure	25

Définition

L'Analyse Discriminante Bayésienne (ADB) permet de construire à partir d'un ensemble de variables quantitatives et d'une variable qualitative découpant la population en plusieurs groupes (2 ou plus), des fonctions discriminantes qui définissent une règle de décision optimale à partir de laquelle on peut affecter des individus de validation et de prévision aux différents groupes.

Cette technique suppose que l'on connaisse a priori les probabilités d'appartenance aux différents groupes et que les données suivent une loi multi-normale.

La méthode proposée permet de traiter les cas linéaire (égalité des matrices de variances) et quadratique (non-égalité des matrices de variances).

L'entrée des probabilités a priori est proposée. Par défaut, le système utilise les probabilités issues des fréquences des groupes dans les données entrées.

En fonction des données et des paramètres définis par l'utilisateur, l'analyse ADB réalise automatiquement les études de la population d'apprentissage et des éventuelles populations de validation et de prévision.

De façon plus précise, la méthode peut se décomposer en trois étapes. Supposons une population de n individus. Découpons cette population en trois sous-populations de tailles n_1 , n_2 et n_3 avec $n_1 + n_2 + n_3 = n$. Les trois étapes sont :

- une étude initiale sur la population d'apprentissage de taille n1
- une étude de validation sur la population de validation de taille n2
- une étude prospective sur une population de prévision de taille n3

Des tableaux résumés et détaillés des classements sont calculés. Le tracé de plans factoriels et un rapport général de synthèse sont proposés.

Entrée des données

Cliquons sur l'icône ADB dans le ruban Expliquer. La boîte de dialogue montrée ciaprès s'affiche :

Analyse discriminante bayésienne			\times
^	▶	Facteur de classement:]
	\square	Variables explicatives quantitatives:	
		(Libellés des variables explicatives:)]
		(Libellés des individus:)]
		(Probabilités initiales:)]
~		Centrage et réduction Oui	
Ok Annuler	Sélectio	on Supprimer Aide	

Cette boîte de dialogue permet de préciser la variable qualitative définissant facteur de classement, les variables explicatives quantitatives, la variable contenant les libellés des variables explicatives et la variable contenant les libellés des individus.

Elle permet également de définir la variable contenant les probabilités a priori si on ne désire pas qu'UNIWIN les calcule à partir des fréquences des groupes dans les données.

Enfin, l'option de centrage et réduction des données est proposée. Cette option est utile si le classement de certains individus n'est pas possible suite à un dépassement de capacité lors des calculs. Il se peut que cela soit lié aux grandeurs des données et les centrer-réduire peut résoudre ce problème ce calcul.

Données manquantes

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

Exemple 1 : Fichier IRIS3

Nous utiliserons le fichier IRIS3 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 et largeur du pétale

Les trois espèces sont : Iris Setosa, Iris Versicolor et Iris Virginica

Cliquons sur l'icône ADB dans le ruban Expliquer.

La boîte de dialogue montrée ci-après s'affiche.

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

Réalisons une analyse non centrée-réduite et laissons à UNIWIN le soin de calculer les probabilités initiales.

Cliquons sur le bouton Sélection pour définir la population d'apprentissage.

		Factor y de classement:
ype onsepal arsepal onpetal		codesp2
arpetal codesp1 codesp2 numiris		Variables explicatives quantitatives:
nesures nomesp		lonpetal larpetal
		~
		(Libellés des variables explicatives:) mesures
		(Libellés des individus:)
		(Probabilités initiales:)
		Centrage et réduction Oui Non
Ok Annuler	 ✓ Séle 	ction Supprimer Aide

Définitio	on de la sélection				×
Et	type		=	A	
Liaison	Variable		Relation	Valeur ou variable	
Et Et non Ou Ou non	larsepal lonpetal lonsepal mesures nomesp numiris type	~	= <> <= > >= débute	codesp1 codesp2 larpetal larsepal lonpetal lonsepal mesures	* *
	Ok	Annuler	A	ijouter	Aide

Définition de la sélection	×
La sélection définie comporte 120 éléments.	
ОК	

Cette sélection comporte 120 individus qui constituent la population d'apprentissage.

Les individus non sélectionnés pour lesquels les valeurs du facteur de classement sont connues constituent la population de validation.

Les individus non sélectionnés pour lesquels les valeurs du facteur de classement ne sont pas connues constituent la population de prévision.

Après avoir renseigné cette boîte de dialogue, UNIWIN débute le calcul de l'Analyse Discriminante Bayésienne.

Après quelques instants, un tableau précisant l'inertie expliquée par les différents vecteurs propres issus de l'analyse apparaît ainsi qu'un diagramme des pourcentages d'inertie expliquée par chacun des axes.



L'option 'Nombre d'axes à retenir' permet de préciser le nombre de composantes principales à extraire.

Cliquons sur le bouton 'Suite des calculs'.

Après quelques instants, UNIWIN affiche une boîte de dialogue vous indiquant si l'hypothèse d'égalité des matrices de variances est vérifiée ou non.



Dans notre exemple, l'hypothèse doit être rejetée et donc une analyse discriminante quadratique est suggérée.

Choisissons cependant de mettre en œuvre une analyse linéaire.

Après quelques instants, l'écran suivant s'affiche :

Rapports et Graphiques									<u>_ 🗆 ×</u>		
Rapport ADB		1	2	3	4	5	6	7	8		
- Probabilités initiales	1										
Centroi des	2	(C) UNIWIN version 9.7.0									
- Corrélations globales	3										
Covariances par groupe	4	DATE: 26/07/2023									
E - Corrélations par groupe	5	ORDINATEUR : LAPTOP-LEG8L077									
Covariances intra-groupes	- Covariances intra groupes 6 UTILISATEUR : cohar										
Correlations intra-groupes Techner(s) DE DONNEES OUVERT(S): IRIS3.SGD											
Tableau des inerties	Tableau de instruites 8										
Test de Pillai	9	RESULTATS DE L'ANALYSE DISCRIM			I						
Test de Box	10										
Fct, discriminante std.	11	Sélection :									
Résultats variables	12	Et type = A									
Résultats individus	13										
- Coord. centres des groupes	14	Méthode : linéaire									
Coer. fct clas. In. (apprentissage) E- Détails classement (apprentissage)	15	Données centrées-réduites : non									
Classement apprentissage	16	Probabilités initiales: 0,35833 0,3									
 Stats - Groupes observés 	B Sate - Groupes observés 17										
Stats - Groupes prévus	18	Jeu d'apprentissage : 120 observat	ions								
Details classement (validation)	19	Jeu de validation : 24 observations									
Classement prévision	20	Jeu de prévision : 6 observations									
	21										
		N									
	7	Rapport Explorateur /									

La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données' Les de rappeler la boîte de dialogue d'entrée des données.

L'icône 'Rapports' affiche la boîte de dialogue des options pour les rapports :

Rapports
Rapport Explorateur
O Rapport Général
O Rapport Html
Ok Appuler
Aunder

1	

et l'icône 'Graphiques' is affiche la boîte de dialogue, montrée ci-après, des options pour les graphiques :

Graphiques
O Diagramme des inerties
O Cercle factoriel des corrélations des variables (points)
O Cercle factoriel des corrélations des variables (points+lignes)
O Plan factoriel du jeu d'apprentissage
◯ Plan factoriel du jeu de validation
O Plan factoriel du jeu de prévision
O Plan factoriel des jeux d'apprentissage et de validation
O Plan factoriel des jeux d'apprentissage et de prévision
◯ Courbe ROC pour le jeu d'aprentissage
◯ Courbe ROC pour le jeu de validation
Ok Annuler

L'icône 'Enregistrer' permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.

Enregistrement des résultats (1/2)	
Enregistrer	Noms attribués aux variables cibles
Fct. discri. std. (apprentissage)	fdstdA_1
Fct. discri. non std. (apprentissage)	fdnstdA_1
Coordonnées individus (apprentissage)	cindA_1
Libellés individus (apprentissage)	libindA
Classes individus (apprentissage)	clindA
Coordonnées individus (validation)	cindV_1
Libellés individus (validation)	libindV
Classes individus (validation)	clindV
Distances carrées individus (apprentissage)	distindA
Cosinus carrés individus (apprentissage)	cosindA_1
Ok Plus	Tout Annuler

Note : le bouton 'Plus' permet d'afficher la suite de la liste des variables.

L'icône 'Quitter' permet de quitter l'analyse.

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.

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

										_		
Rapports et Graphiques									_			
	s											
Rapport ADB		1	2	3	4	5	6	7	8	^		
Probabilités initiales	1											
Centroï des	2	RESULTATS INDIVIDUS POUR LE FAC	TFUR:1									
Covariances globales	-											
Corrélations globales				DUCCUTOS								
E Covanances par groupe	4	DISTANCE^2 = CARRES DES DISTANC	ES A L'ORIGINE OU AU BA	ARYCENTRE						_		
Council and a service and a se	5	COORD. = COORDONNEES DES INDIVID	DUS							-		
Consistions intra-groupes	6	CONTRIB. = CONTRIBUTIONS A L'INEF	RTIE									
Distances de Mabalanobis	7	COSINUS ² = COSINUS CARRES	SINUS^2 = COSINUS CARRES									
Tableau des inerties	8	COS^2 CUM. = SOMMES CUMULEES	DES COSINUS CARRES									
Test de Pillai	9											
Test de Box	10											
Fct. discriminante std.	44		CROUPE	ND5/DU	DISTANCEAD	00000	CONTRIP	CORINIERO	CO542 CUM			
Fct. discriminante non std.			GROUPE	INDIVIDU	DISTANCE'2	COORD.	CONTRO.	CUSINU3-2	CO3-2 COM.			
Résultats variables	12	1	1	1	57,32730	7,56828	1,48724	0,99916	0,99916	-		
E- Resultats individus	13	2	1	2	45,52882	6,69560	1,16403	0,98467	0,98467	_		
Eactour 2	14	4	1	3	40,65299	6,34050	1,04384	0,98890	0,98890			
- Points multiples individus	15	5	1	4	58,20534	7,61605	1,50607	0,99655	0,99655			
Coord, centres des groupes	16	6	1	5	53,47025	7,18429	1,34015	0,96528	0,96528			
Coef. fct clas. lin. (app)	17	7	1	6	45.09110	6.70541	1,16744	0.99715	0.99715			
Détails classement (app)	18	8	1	7	50 66633	7 11765	1 31540	0.99989	0.99989			
Classement apprentissage	40	9		,	20,00000	6,40453	0.06750	0,03000	0,03663			
Image: Stats - Groupes observés	19	9	1	0	30,24330	6,10453	0,90759	0,97442	0,97442	-		
Stats - Groupes prévus	20	10	1	9	62,60347	7,89480	1,61833	0,99560	0,99560	-		
Details classement (valid)	21	11	1	10	45,10741	6,71478	1,17071	0,99958	0,99958			
Classement prévision	<	Rapport Explorateur /	1	**	40 50505	0 07000	* 22204	0.07/07	0.07407	>		

Rap	Rapports et Graphiques													
	2 🐺 💵 💈	l												
	1	2	3	4	5	6	7	8	9	10	11	12	13	^
494														
495	ANALYSE LINEAIRE - COEFFI	CIENTS DES FONCT	IONS DE CLASSEME	NT										
496														
497														
498		Setosa	Versicolor	Virginica										
499	Constante	-85,34195	-72,20247	-103,06840										
500	lonsepal	22,72256	14,66467	10,78294										
501	larsepal	23,46884	8,91552	5,92113										
502	lonpetal	-14,98624	5,89627	13,90215										
503	larpetal	-17,29620	4,45679	19,29284										
504														
505	DISCRIMINATION LINEAIRE													
506														
507	GROUPE OBSERVE : Setosa													
508														
509	JEU D'APPRENTISSAGE													
510														
511	L'INDIVIDU EST AFFECTE AU	GROUPE DE PLUS F	ORTE PROBABILITE											
512														
513														
514	INDIVIDU-GROUPE	P(Setosa)	P(Versicolor)	P(Virginica)										
<>	Rapport Général /	<	0.00000	0 00000										>

E	Rapports et G	iraphiques						x
			1					
L	i148 - Virgini	ica	0,00000	0,0019	98 0,	99802		^
	ANALYSE LIN	IEAIRE - R	ESULTATS D	U CLASSEM	ENT POUR	LE JEU D'A	PPRENTISSAGE	
	EN LIGNES, L EN COLONNE	ES GROU	PES OBSERN	/ES VUS				
	POURCENTA	GE DE MA GE DE BIE	L CLASSES : N CLASSES :	1,667 % 98,333 %				
L		Setosa	Versicolor	Virginica	Total			
L	Setosa	43	0	0	43			
L	Versicolor	0	37	1	38			
L	Virginica	0	1	38	39			
L	Total	43	38	39	120			
	STATISTIQUE	S GROUF	E OBSERVE :	Setosa				
L								
			lonsepal	larsepal	lonpetal	larpetal		
	Effectif		43,00000	43,00000	43,00000	43,00000		
L	Moyenne		5,02326	3,44419	1,46512	0,25349		~

Ces rapports nous fournissent les renseignements suivants :

- Probabilités initiales
- Centroïdes des groupes et global
- Matrices des covariances et corrélations globales et des groupes
- Matrice des covariances et des corrélations intra-groupes
- Distances de Mahalanobis entre les groupes, Fishers, niveaux de signification
- Tableau des inerties (avec corrélation canonique, lambda de Wilks, Khi-2, degrés de liberté et niveau de signification
- Test de Pillai
- Test de Box
- Fonctions discriminantes standardisées et non standardisées
- Résultats pour les variables et pour les individus
- Coordonnées des centres des groupes
- Coefficients des fonctions de classement (analyses linéaire et/ou quadratique)
- Résumé du classement d'apprentissage (analyses linéaire et/ou quadratique)
- Tableaux détaillés des classements par groupe (apprentissage)
- Statistiques pour les groupes observés et prévus (apprentissage)
- Résumé et détails du classement de la population de validation
- Classement de la population de prévision

L'option Graphiques

• Diagramme des inerties

Ce graphique affiche les pourcentages d'inertie pour chacun des axes factoriels.



Cercle factoriel des corrélations des variables

Ces options permettent d'afficher le cercle de corrélations des variables et de choisir si on désire tracer les lignes reliant les points à l'origine du cercle. L'option sans ces lignes est utile lorsqu'il y a un grand nombre de variables représentées.

Choisissons les variables avec lignes puis sans lignes. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet également de préciser si l'on désire afficher les libellés des variables, de choisir la couleur et la police et d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.

Axe horizontal	Axe vertical		Réutiliser les titres
2	1	A.	 Non Oui
bellés) Sans) Avec) Groupes observés	îmes New Ron Normal	12	Police Couleur
) Groupes prévus			





• Plan factoriel des individus et centres des groupes

Ces options permettent d'afficher des plans factoriels des individus et des centres des groupes pour les populations d'apprentissage, de validation et de prévision. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche.

Elle permet de préciser si l'on désire afficher ou non les libellés des individus, de préciser si ces libellés sont les codes des groupes observés ou les codes des groupes prévus, de choisir la couleur et la police pour ces libellés. Il est également possible d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.

Choix du plan factorie	ł			×
Axe horizontal	Axe vertical		Réutiliser les titres	
1 2	▲ 1 2	* *	● Non ○ Oui	
Libellés Sans Avec Groupes observés Groupes prévus	Times New Ron Normal	12	Police Couleur	
Défaut	Ok	Sélection	Annuler	

Des exemples de plans factoriels sont montrés ci-après.













Courbe ROC

Ces deux options ne sont pas disponibles dans cette analyse car la variable à expliquer a plus de deux modalités.

Exemple 2 : Fichier INFARCT2

Pour ce deuxième exemple, nous utiliserons le fichier INFARCT2.

Ce fichier contient des informations concernant 101 victimes d'un infarctus du myocarde. Cette population est constituée d'une population d'apprentissage de 81 individus et d'une population de validation de 20 individus.

La population d'apprentissage comprend les individus 11 à 51 (groupe Décès) et les individus 52 à 91 (groupe Survie).

La population de validation comprend les individus 1 à 10 (groupe Décès) et les individus 92 à 101 (groupe Survie).

Les variables mesurées sont :

Variable	Mesure
frcar incar insys prdia papul pvent	fréquence cardiaque index cardiaque index systolique pression diastolique pression artérielle pulmonaire pression ventriculaire
repul	résistance pulmonaire

La variable *groupe* indique le groupe d'appartenance de chaque individu (Décès ou Survie). La variable *type* précise la population d'appartenance de chaque individu (A si apprentissage, V si validation). La variable *libobs* contient les libellés des individus des populations d'apprentissage et de validation.

Cliquons sur l'icône ADB dans le ruban Expliquer. La boîte de dialogue montrée cidessous s'affiche.

Analyse discriminante bayésienne		×
frcar incar insys prdia papul pvent repul groupe type libobs		Facteur de classement: groupe Variables explicatives quantitatives: frcar incar insys prdia papul pvent repul
		(Libellés des variables explicatives:) (Libellés des individus:) libobs
		(Probabilités initiales:) Centrage et réduction Oui Non
Ok Annuler	Sélectio	n Supprimer Aide

Cliquons sur le bouton Sélection pour définir la population d'apprentissage.

^D Définitio	on de la sélection					×
Et	type		=	A		
Liaison	Variable		Relation	Valeur ou variable		
Et Et non Ou Ou non	insys libobs papul prdia pvent repul type	~	= <> <= > >= débute	frcar groupe incar insys libobs papul prdia		^
	Ok	Annuler	Aj	outer	Aide	

Un message nous indique que 81 individus sont sélectionnés.

Puisque la variable à expliquer possède deux classes, la procédure demande de préciser le code de l'événement positif qui sera utilisé pour le tracé de la courbe ROC.

Options pour la courbe ROC	×
Code de l'événement positif pour l'analyse	
O Décès O Survie	
Ok Annuler	

Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse.

Après quelques instants, un tableau précisant l'inertie expliquée par l'unique vecteur propre issu de l'analyse s'affiche.

T 11	Tableau des Inetties										
Nom	Nombre d'axes à retenir : 1 v Suite des calculs										
	1	2	3	4	5	6	7	8	9	10	^
1											
2	Tableau des Inert	ies									
3											
4	Composante	Valeur propre	Pct de variance	Pct cumulé	Variation	Corrél. canon.					
5	1	1,34960	100	100	0	0,75789					
6											

Aucun graphique n'est proposé, car il n'y a qu'une unique valeur propre.

Après quelques instants, UNIWIN affiche une boîte de dialogue vous indiquant si l'hypothèse d'égalité des matrices de variances est vérifiée ou non.



Dans notre exemple, l'hypothèse doit être rejetée et donc une analyse discriminante quadratique est suggérée. Choisissons donc de mettre en œuvre une analyse quadratique.

Après quelques instants, l'écran suivant s'affiche :

Rapports et Graphiques									<u>_ 🗆 ×</u>
Rapport ADB		1	2	3	4	5	6	7	8
 Probabilités initiales 	1								
Centroï des	2	(C) UNIWIN version 9.7.0							
Covilations globales	3								
Coveriances par groupe	4	DATE : 11/08/2023							
Corrélations par groupe	-	OPDINATEUR : LAPTOR LECALO77							
- Covariances intra-groupes		URDINATEUR : LAPTOP-LEGOLOTT							
Corrélations intra-groupes	0	UTILISATEUR : cchar							
- Distances de Mahalanobis	7	FICHIER(S) DE DONNEES OUVERT(S) :	INFARCT2.SGD						
Tablesu des ineties 8									
Test de Pillai	9	RESULTATS DE L'ANALYSE DISCRIMI	NANTE BAYESIENNE						·
- Ect. discriminante std	- Tert doe Box								
Fct, discriminante non std.	11	Sélection :							
Résultats variables	12	Et type = A							
B- Résultats individus	13								
Coord. centres des groupes	14	Méthode : guadratique							
Loef. tct. clas. quad. (apprentissage)	15	Données centrées-réduites : non							
Classement apprentissage	16	Probabilités initiales : 0.50617 0.4	9383						
VP, FN, FP, VN, Sensibilité, Spécificité	17								
Stats - Groupes observés	40	In discontinuos of abananti							
Stats - Groupes prévus	10	Jeu d'apprentissage : et observation	лів						
Détails classement (validation)	19	Jeu de validation : 20 observations							
Classement validation	20	Jeu de prévision : 0 observations							
···· vr, riv, rr, viv, Sensibilite, Specificite (21								
•		Rapport Explorateur	7						

La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données' de rappeler la boîte de dialogue d'entrée des données, par l'icône 'Rapports' d'afficher la boîte de dialogue des options pour les rapports et par et par l'icône 'Graphiques' d'afficher la boîte de dialogue des options pour les graphiques.

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.

Ces rapports nous fournissent les renseignements suivants :

- Probabilités initiales
- Centroïdes des groupes et global
- Matrices des covariances et corrélations globales et des groupes
- Matrice des covariances et des corrélations intra-groupes

- Distances de Mahalanobis entre les groupes, Fishers, niveaux de signification
- Tableau des inerties (avec corrélation canonique, lambda de Wilks, Khi-2, degrés de liberté et niveau de signification
- Test de Pillai
- Test de Box
- Fonctions discriminantes standardisées et non standardisées
- Résultats pour les variables et pour les individus
- Coordonnées des centres des groupes
- Coefficients des fonctions de classement (analyses linéaire et/ou quadratique)
- Résumé du classement pour la population d'apprentissage (analyses linéaire et/ou quadratique)
- Tableaux détaillés des classements par groupe (apprentissage)
- VP, FN, FP, VN, sensibilité et spécificité pour la population d'apprentissage
- Statistiques pour les groupes observés et prévus (apprentissage)
- Résumé et détails du classement de la population de validation
- VP, FN, FP, VN, sensibilité et spécificité pour la population de validation
- Classement de la population de prévision (si existe)

Rapports et Graphiques									_ 🗆 X
]								
Rapport ADB		1	2	3	4	5	6	7	8
 Probabilités initiales 	1								
Centroi des	2	RESULTATS INDIVIDUS POUR LE FACT	TEUR : 1						
Covanances globales	3								
Covariances par groupe	4	DISTANCE^2 = CARRES DES DISTANC	ES A L'ORIGINE OU AU B	ARYCENTRE					
E - Corrélations par groupe	5	COORD. = COORDONNEES DES INDIVID	US						
Covariances intra-groupes	6	CONTRIB. = CONTRIBUTIONS A L'INER	TIF						
Corrélations intra-groupes	7	COSINUS^2 = COSINUS CARRES							
Distances de Manaianobis	8	COSA2 CIIM - SOMMES CIIMIU EES D							
Test de Pilai			LI COMINI CARLES						
Test de Box	- 40								
Fct. discriminante std.	10								
- Fct. discriminante non std.	11		GROUPE	INDIVIDU	DISTANCE"2	COORD.	CONTRIB.	COSINUS*2	COS^2 CUN
Résultats variables	12	11	1	1	0,12121	-0,34815	0,06531	1	
Hésultats individus	13	12	1	2	2,88426	1,69831	1,55397	1	
- Coord centres des groupes	14	13	1	3	0,00002	-0,00455	0,00001	1	
⊕-Coef. fct. clas. guad. (apprentissage)	15	14	1	4	4,83268	2,19833	2,60372	1	
Détails classement (apprentissage)	16	15	1	5	1,90155	1,37897	1,02451	1	
Classement apprentissage	17	16	1	6	2,85242	1,68891	1,53681	1	
- VP, FN, FP, VN, Sensibilité, Spécific	18	17	1	7	1,29329	1,13723	0,69679	1	
Stats - Groupes observes	19	18	1	8	0.22116	0.47028	0.11915	1	
	20	19	1	9	3.80729	1,95123	2.05127	1	
Classement validation	21	110	1	10	1 19438	1,00728	0.64350	1	
VP, FN, FP, VN, Sensibilité, Spécific		214		10	1,19430	1,03200	0,04330	1	
		Rapport Explorateur /							

Rapports et Graphiques									_ 🗆 X
Rapport ADB		1	2	3	4	5	6	7	8
 Probabilités initiales 	1								
Centroï des	2	ANALYSE QUADRATIQUE - RESULTAT	S DU CLASSEMENT POU	R LE JEU D'APPRENTISS	AGE				
Covariances globales	3								
Covariances par groupe	4	EN LIGNES, GROUPES OBSERVES							
Corrélations par groupe	5	EN COLONNES GROUPES PREVIIS							
Covariances intra-groupes	6	EN COLONNES, SNOUPES PREVUS							
- Corrélations intra-groupes									
- Distances de Mahalanobis		POURCENTAGE DE MAL CLASSES : 1	1,111 %						
Tableau des inerties	8	POURCENTAGE DE BIEN CLASSES : 84	3,889 %						
Test de Pila	9								
- Ect_discriminante std	10								
Fct, discriminante non std.	11		Décès	Survie	Total				
B- Résultats variables	12	Décès	36	5	41				
Résultats individus	13	Survie	4	36	40				
Coord. centres des groupes	14	Total	40	41	81				
Détaile classement (apprentissage)	15								
Classement apprentissage	16								
- VP, FN, FP, VN, Sensibilité, Spécificité	17								
 Stats - Groupes observés 	19								
Interpretation - Groupes prévus	10								
Détails classement (validation)	19								
	20								
···· vr, riv, rr, viv, sensibilite, specificite i	21								
		Rapport Explorateur /					1		

L'option Graphiques

• Diagramme des inerties

Ce diagramme n'est pas disponible dans cet exemple car il n'y a qu'une seule composante.

• Cercle factoriel des corrélations des variables

Ces options permettent d'afficher le cercle de corrélations des variables et de choisir si on désire tracer les lignes reliant les points à l'origine du cercle. L'option sans ces lignes est utile lorsqu'il y a un grand nombre de variables représentées. Choisissons les variables avec lignes puis sans lignes. Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet également de préciser si l'on désire afficher les libellés des variables, de choisir la couleur et la police et d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.

Choix du plan factorie	:		×
Axe horizontal	Axe vertical		Réutiliser les titres
1	1	×	 Non Oui
Libellés			
Groupes observés Groupes prévus	Times New Ron Normal	12	Police Couleur
Défaut	Ok	Sélection	Annuler

A noter que dans notre exemple, il n'y a qu'un axe factoriel et donc le tracé n'est pas proposé.

• Plan factoriel des individus et centres des groupes

Ces options permettent d'afficher des plans factoriels des individus et des centres des groupes pour les populations d'apprentissage, de validation et de prévision.

Une boîte de dialogue permettant de choisir le plan factoriel s'affiche. Elle permet de préciser si l'on désire afficher ou non les libellés des individus, de préciser si ces libellés sont les codes des groupes observés ou les codes des groupes prévus, de choisir la couleur et la police pour ces libellés. Il est également possible d'indiquer si les titres du graphique (titre 1, titre 2), doivent être conservés pour être réutilisés ultérieurement dans d'autres graphiques créés lors de cette même session de travail.

Axe horizontal	Axe vertical		Réutiliser les titres
		A 	● Non ○ Oui
ibellés ● <u>Sans</u> ● Avec ● Groupes observés ^T ● Groupes prévus	ìmes New Ron	12	Police Couleur



Refaisons maintenant cette même analyse en choisissant une méthode linéaire.

Pour cela, il faut refaire l'analyse actuelle en choisissant 'Non' lorsque la boîte de dialogue 'Test de Box' s'affiche.

Rapports et Graphiques									_ 🗆 ×
Rapport ADB		1	2	3	4	5	6	7	8
 Probabilités initiales 	1								
Centroi des	2	ANALYSE LINEAIRE - RESULTATS DU	CLASSEMENT POUR LE J	EU D'APPRENTISSAGE					
Covenances globales	3								
Covariances par groupe	4	EN LIGNES, LES GROUPES OBSERVES							
 Corrélations par groupe 	5	EN COLONNES, LES GROUPES PREVUS	5						
Covariances intra-groupes	6								
Distances de Mahalanchia	7	POURCENTAGE DE MAL CLASSES :	.877 %						
Tableau des inerties	8	POURCENTAGE DE BIEN CLASSES : 90	.123 %						
- Test de Pilai	9		.,						I
Test de Box	10								
- Fct. discriminante std.	11		Dácàs	Survia	Total				
	12	Dácàs	20	2	10101				
⊕ Résultats valables	12	Survie	50	35	41				
Coord. centres des groupes	13	Total	3	33	40				
Coef. fct clas. lin. (apprentissage)	45	Total	40	30	01				
Détails classement (apprentissage)	10								
VP EN EP VN Sensibilité Spécificité	10								
Stats - Groupes observés	17								
Stats - Groupes prévus	18								
Détails classement (validation)	19								
VP EN EP VN Sepaibilité Spécificité	20								
www.r., r.v., r.v., sensibilite, specificite i	21								
	<u> </u>	Rapport Explorateur /							

Le pourcentage d'erreur de classement (9,88 %) est meilleur que pour l'analyse quadratique (11,11 %) mais l'hypothèse d'égalité des matrices des variances, non observée, y est supposée.

• Courbe ROC

Le tracé de la courbe ROC et le calcul de l'aire sous la courbe (AUC) sont possibles car la variable à expliquer possède deux modalités.

Visualisons les courbes pour le jeu d'apprentissage et le jeu de valdiation.



Exemple 3 : Fichier BORDEAUX

Pour ce troisième exemple, nous utiliserons le fichier BORDEAUX.

Ce fichier contient des informations sur la qualité de vins de Bordeaux en relation avec les conditions météorologiques.

La variable qualitative Qualité, facteur de classement, prend trois modalités : Bon, Moyen et Médiocre.

Les variables explicatives sont :

- *Temp* somme des températures moyennes journalières (° C)
- Insol
 durée d'insolation (heures)
- Chaleur nombre de jours de grande chaleur
- Pluie hauteur des pluies (millimètres)

La variable Année fournit les libellés des individus.

Cliquons sur l'icône ADB dans le ruban Expliquer. La boîte de dialogue montrée cidessous s'affiche :

Analyse discriminante bayésienne		×
Année Temp Insol Chaleur Pluie Qualité		Facteur de classement: Qualité Variables explicatives quantitatives: Temp Insol Chaleur Pluie
		(Libellés des variables explicatives:) (Libellés des individus:) Année (Probabilités initiales:) Centrage et réduction ©Qui ONon
Ok Annuler	Sélectio	on Supprimer Aide

Nous sélectionnons une analyse centrée et réduite.

La fenêtre 'Rapports et Graphiques' s'affiche pour la méthode linéaire.

Visualisons les résultats du classement.

Les vins mal classés sont les vins des années :

1953, 1955bons classés moyens1933, 1950moyens classés bons1926moyen classé médiocre1935, 1956médiocres classés moyens

Rapports et Graphiques									_ 🗆 ×
Rapport ADB		1	2	3	4	5	6	7	8
Probabilités initiales	1								
Centroi des	2	ANALYSE LINEAIRE - RESULTATS DU	CLASSEMENT POUR LE	JEU D'APPRENTISSAGE					
- Corrélations globales	3								
Covariances par groupe	4	EN LIGNES, LES GROUPES OBSERVES							
 Corrélations par groupe 	5	EN COLONNES, LES GROUPES PREVUS	5						
Covariances intra-groupes	6								
Correlations intra-groupes	7	POURCENTAGE DE MAL CLASSES : 2	0,588 %						
- Tableau des inerties	8	POURCENTAGE DE BIEN CLASSES : 7	,412 %						
Test de Pilai	9								1
Test de Box	10								
Fct, discriminante std.	11		Bon	Moyen	Médiocre	Total			
 Résultats variables 	12	Bon	9	2	0	11			
⊕- Résultats individus	13	Moyen	2	8	1	11			
Coord. centres des groupes	14	Médiocre	0	2	10	12			
Coef. tct clas. lin. (apprentissage)	15	Total	11	12	11	34			
Classement apprentissage	16								
B- Stats - Groupes observés	17								
Stats - Groupes prévus Stats - Groupes Stats Stats - Groupes Stats Stats	18								
	19								
	20								
	21								
		· · · ·							
	1	Rapport Explorateur /							

W Rapports et Graphiques										_ ×
	A									
Rapport ADB		1	2	3	4	5	6	7	8	^
Probabilités initiales	1									
- Centroi des	2	DISCRIMINATION LINEAIRE								
Covanances globales	3									
+- Covariances par groupe	4	GROUPE OBSERVE : Bon								
Corrélations par groupe	5									
- Covariances intra-groupes	6	JEU D'APPRENTISSAGE								
Corrélations intra-groupes	7									
	8	L'INDIVIDUEST AFFECTE AU GROUPE	DE PLUS FORTE PROBAR							
Test de Pilai	-		DET EUOTONIET NODING							
Test de Box	10									
Fct. discriminante std.	11		P(Bop)	P(Moyan)	D/Médiocra)					
Fct. discriminante non std.	12	1928 - Bon	0.64171	0.32703	0.03126					
Bésultats individus	12	1929 - Bon	0,04111	0,02703	0,00120					
Coord. centres des groupes	13	1924 Bon	0,00044	0,00000	0,00003					
Coef. fct clas. lin. (app)	14	1037 Bee	0,09244	0,10745	0,00012					
Détails classement (app)	15	1937 - Boli	0,62214	0,37672	0,00114					
Mauro	16	1943 - Bon	0,90475	0,09520	0,00004					
Médiocre	17	1945 - Bon	0,98380	0,01620	0,00000					
Classement apprentissage	18	1947 - BON	0,99656	0,00344	0,00000					_
Stats - Groupes observés	19	1949 - Bon	0,99904	0,00096	0,00000					
Bon	20	1952 - Bon	0,85207	0,14781	0,00012					
Moyen	21	1953 - Moyen *	0,19487	0,75191	0,05322					
+- Stats - Groupes prévus	22	1955 - Moyen *	0,38292	0,60530	0,01178					
	23									~
	< > \	Rapport Explorateur / <								>

Demandons également le cercle factoriel et le plan factoriel pour les axes 1 et 2.



Exemple 4 : Fichier TITANIC

Pour ce quatrième exemple, nous utiliserons le fichier TITANIC.

Ce fichier contient des informations concernant 714 passagers :

Statut	Survie ou Décès
Classe	Classe du passager (1 ^{ère} , 2 ^{ème} ou 3 ^{ème})
Sexe	Homme ou Femme
Age	Age du passager
Nbfse	Nombre de frères, sœurs ou époux, épouses à bord
Nbpe	Nombre de parents ou enfants à bord
Tarif	Tarif passager (en £)

Cliquons sur l'icône ADB dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-dessous.

Analyse discriminante bayésienne		×
Statut Age Tarif Nbfse Nbpe Classe Sexe Poids LibVarQuanti LibVarQuali LibObs		Facteur de classement : Statut Variables explicatives quantitatives : Age Tarif Nbfse Nbpe
		(Libellés des variables explicatives :) (Libellés des individus :)
Ok Annuler	Sélectio	(Probabilités initiales :) Centrage et réduction Ouu Non Supprimer Aide

Après exécution de la procédure, visualisons le tableau de classement des données d'apprentissage et la courbe ROC associée.

Rapports et Graphiques									_ 🗆 🗙
Rapport ADB		1	2	3	4	5	6	7	8
Probabilités initiales	1								
- Centroi des	2	ANALYSE QUADRATIQUE - RESULTAT	S DU CLASSEMENT POU	R LE JEU D'APPRENTISS	AGE				
Covanances globales	3								
Covariances par groupe	4	EN LIGNES, GROUPES OBSERVES							
 Corrélations par groupe 	5	EN COLONNES, GROUPES PREVUS							
Covariances intra-groupes	6								
Corrélations intra-groupes	7	POURCENTAGE DE MAL CLASSES : 3	0.952 %						
Tableau des inerties	. 8	POURCENTAGE DE BIEN CLASSES : 69							
- Test de Pillai			,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,						I
Test de Box	10								
Fct. discriminante std.	10		0(-)-	Currie	Total				
Fct. discriminante non std.		Díala	Deces	Survie	Total				
Résultats vanables	12	Deces	402	22	424				
Coord, centres des groupes	13	Survie	199	91	290				
Coef. fct. clas. quad. (apprentissage)	14	iotai	601	113	714				
Détails classement (apprentissage)	15								
	16								
	17								
Stats - Groupes prévus	18								
	19								
	20								
	21								
•		Rapport Explorateur /							

Environ 69 % des passagers sont bien classés par cette analyse et l'aire sous la courbe ROC est proche de 0,7.

<u>Note</u> : Pour comparer les performances de plusieurs méthodes d'analyse, cet exemple est traité dans les six analyses AFD, ADB, KNN, BAYES, ANN et ARBRE.

Les variables internes créées par la procédure

Voici la liste des variables internes créées par la procédure. Ces variables peuvent notamment être utilisées avec l'option 'Sélection'. A noter que certaines des variables mentionnées ci-dessous peuvent ne pas apparaître, en fonction des options choisies.

Variable	Contenu
fdstdA	Fonctions discriminantes standardisées (apprentissage)
fdnstdA	Fonctions discriminantes non standardisées (apprentissage)
cindA	Coordonnées des individus (apprentissage)
libindA	Libellés des individus (apprentissage)
clindA	Classes des individus (apprentissage)
cindV	Coordonnées des individus (validation)
libindV	Libellés des individus (validation)
clindV	Classes des individu (validation)
distindA	Distances carrées à l'origine pour les individus (apprentissage)
cosindA	Cosinus carrés pour les individus (apprentissage)
conindA	Contributions pour les individus (apprentissage)
cvarA	Coordonnées des variables (apprentissage)
disvarA	Distances carrées à l'origine des variables (apprentissage)
cosvarA	Cosinus carrés des variables (apprentissage)
convarA	Contributions des variables (apprentissage)

vpA	Vrais positifs (apprentissage)
fnA	Faux négatifs (apprentissage)
fpA	Faux positifs (apprentissage)
vnA	Vrais négatifs (apprentissage)
specificiteA	Spécificité (apprentissage)
sensibiliteA	Sensibilité (apprentissage)
vpV	Vrais positifs (validation)
fnV	Faux négatifs (validation)
fpV	Faux positifs (validation)
vnV	Vrais négatifs (validation)
specificiteV	Spécificité (validation)
sensibiliteV	Sensibilité (validation)

si modèle linéaire :

coefcl	Fonctions de classement linéaire (apprentissage)
classIA	Classement linéaire (apprentissage)
classIV	Classement linéaire (validation)
classIP	Classement linéaire (prévision)
libindP	Libellés des individus (prévision)

si modèle quadratique :

ccq(i)	Constantes classement quadratique groupe i (apprentissage)
coefcq(i)	Fonctions de classement quadratique groupe i (apprentissage)
classqA	Classement quadratique (apprentissage)
classqV	Classement quadratique (validation)
classqP	Classement quadratique (prévision)
libindP	Libellés des individus (prévision)