

UNIWIN VERSION 10.5.0

SCORING

Révision : 26/03/2026

Définition.....	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Fichier Crédit	3
L'option Rapports	8
L'option Graphiques	12
L'option Classement	17
L'option Simulation	18
L'option Enregistrement.....	19
Population de prévision : Fichier Crédit2	20
Exemple 2 : Fichier Ventas.....	22

Définition

Le scoring est une technique permettant de classer des individus caractérisés par des variables descriptives qualitatives dans deux catégories préexistantes en leur affectant un score d'appartenance.

La première étape de cette technique consiste à transformer les variables descriptives qualitatives d'origine en coordonnées sur les axes factoriels issus d'une Analyse des Correspondances Multiples (ACM) des variables qualitatives.

La deuxième étape est une Analyse Factorielle Discriminante (AFD) de ces données, avec une possible sélection des axes factoriels significatifs de l'ACM.

La fonction discriminante obtenue est ensuite exprimée en fonction des indicatrices des modalités des variables qualitatives d'origine et les coefficients de la fonction de score sont évalués.

L'analyse se poursuit en calculant les scores des individus, la sensibilité, la spécificité, les gains cumulés et les valeurs de Lift.

Un rapport général de synthèse est proposé ainsi que des graphiques des coefficients de la fonction de score, des courbes de répartition (avec ou sans zone d'indécision), des courbes de répartition et de densité, de la courbe ROC, de la courbe des gains cumulés et de la courbe Lift.

Une option de simulation permet d'évaluer l'évolution du score d'un individu lorsqu'une ou plusieurs de ses modalités d'origine sont modifiées.

Des tableaux résumés et détaillés des classements de l'Analyse Factorielle Discriminante et du Scoring sont calculés.

L'étude d'une population de prévision peut également être réalisée. Elle évalue les scores de ces nouveaux individus.

Entrée des données

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

Scoring - Apprentissage

Facteur de classement :

Variables explicatives qualitatives :

(Libellés des variables explicatives :)

(Libellés des individus :)

Nombre maximum de facteurs ACM : 100

Sélection des facteurs significatifs

Ok Annuler Sélection Supprimer Aide

Cette boîte de dialogue permet de définir le facteur de classement qualitatif, les variables explicatives qualitatives, la variable contenant les libellés de ces variables explicatives et la variable contenant les libellés des individus.

Le nombre maximum de facteurs à extraire de l'ACM peut être précisé et la case à cocher 'Sélection des facteurs significatifs' permet d'indiquer que seuls les facteurs significatifs devront être utilisés pour définir la fonction de score.

Données manquantes

Les données manquantes ne sont pas autorisées.

Exemple 1 : Fichier Crédit

Nous utiliserons le fichier Crédit pour illustrer cette procédure.

Ce fichier, au format texte, est disponible sur le site « UCI Machine Learning Repository ». Il a été conçu par le Professeur Dr. Hans Hofmann de l'Institut für Statistik und Okonometrie, Universität Hamburg

Il contient un ensemble de 21 informations concernant 1.000 demandeurs d'un crédit. Ces demandeurs sont classés « bons » (700) ou « mauvais » (300). Nous souhaitons construire une fonction de scoring pouvant être utilisée pour déterminer si un nouveau demandeur d'un crédit est un « bon » demandeur (sans risque) ou un « mauvais » demandeur (à risque).

Voici la liste de ces 21 informations :

<i>Variable</i>	<i>Descriptif de la variable qualitative et de ses modalités</i>
Type client	Type de client « Bon client », « Mauvais client »
Compte chèque	Montant sur le compte chèque en DM « 0 DM », « 1 à 200 DM », « Plus de 200 DM » « Pas de compte chèque »
Durée	Durée du crédit en mois « 1 mois », « 2 mois », « 3 mois », « 4 mois »
Historique	Historique des crédits « Pas de crédits », « Crédits terminés payés » « Crédits en cours », « Crédits payés en retard » « Compte critique »

Objet	Objet du crédit « Véhicule neuf », « Véhicule d'occasion » « Biens d'équipement », « Radio ou télévision » « Appareils ménagers », « Réparations » « Education », « Formation », « Travail » « Autres »
Montant	Montant du crédit (variable continue discrétisée par quartiles) « 0 à 1366 DM », « 1367 à 2320 DM » « 2321 à 3973 DM », « 3974 DM et plus »
Compte épargne	Montant sur le compte d'épargne en DM « Moins de 100 DM », « 101 à 500 DM » « 501 à 1000 DM », « Plus de 1000 DM » « Sans épargne »
Ancienneté	Ancienneté dans l'emploi « Sans emploi », « Moins de 1 an » « 1 à 3 ans », 4 = « 4 à 6 ans », « Plus de 6 ans »
Taux	Taux du crédit « 0,01 », « 0,02 », « 0,03 », « 0,04 »
Statut	Statut du demandeur (sexe et civilité) « H divorcé », « F divorcée ou mariée » « H célibataire », « H marié »
Garantie	Garantie, autres débiteurs « Aucune », « Codemandeur » « Caution »
Résident	Nombre d'années de résidence « 1 an », « 2 ans », « 3 ans », « 4 ans »
Biens	Biens possédés « Immobilier », « Epargne, Assurance » « Véhicule », « Sans biens »
Age	Age du demandeur (variable continue discrétisée par quartiles) « Moins de 27 ans », « 27 à 32 ans » « 33 à 41 ans », « Plus de 41 ans »
Autres crédits	Autres crédits en cours « Bancaires », 2 = « Magasins » « Aucun »

Logement	Logement « Locataire », « Propriétaire » « Logement gratuit »
Nombre de crédits	Nombre de crédits en cours « 1 crédit », « 2 crédits », « 3 crédits » « 4 crédits »
Emploi	Type d'emploi « Sans », « Non qualifié » « Employé », « Direction »
Responsable	Nombre de personnes responsables du crédit « Un », « Deux »
Téléphone	Téléphone « Sans », « Avec »
Travailleur	Type de travailleur « Etranger », « National »

Cliquons sur l'icône SCO dans le ruban Expliquer.

La boîte de dialogue montrée ci-après apparaît.

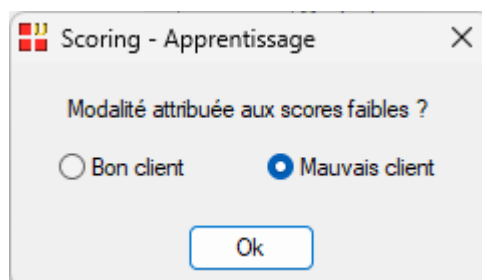
Sélectionnons les variables « Compte chèque » à « Travailleur » comme variables explicatives qualitatives, la variable « Type client » comme facteur de classement et la variable « Individu » comme variable contenant les libellés des individus.

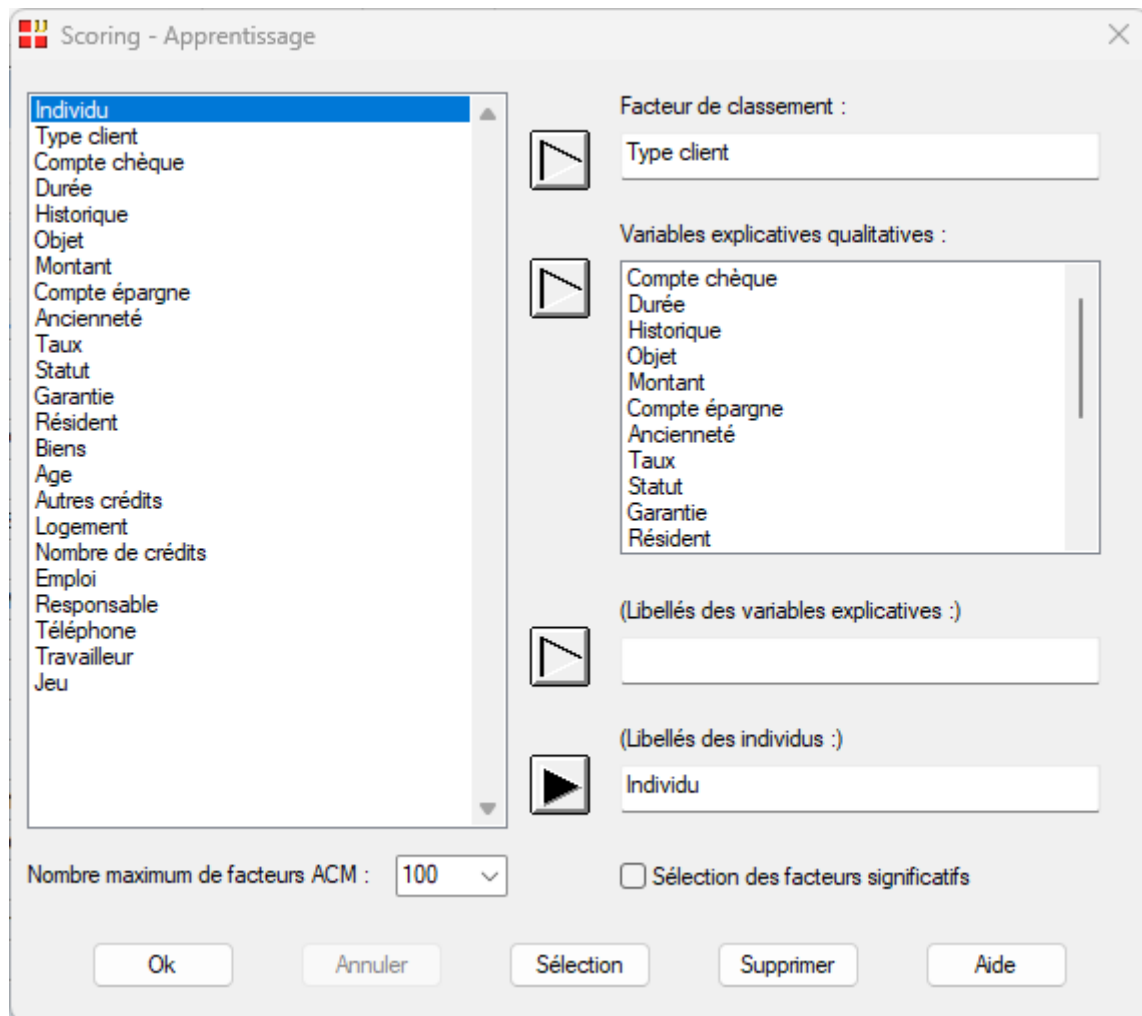
Dans cet exemple, nous ne cocherons pas la case « Sélection des composantes ACM significatives » et accepterons la valeur par défaut pour le nombre maximum de facteurs ACM.

Cliquons sur le bouton Ok.

UNIWIN nous demande alors de préciser la modalité du facteur de classement qui doit être attribuée au groupe des scores faibles.

Indiquons que ce groupe est caractérisé par la modalité « Mauvais client ».







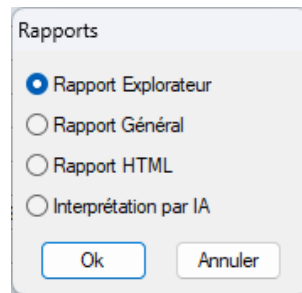
Après avoir renseigné cette boîte de dialogue, UNIWIN débute le calcul du scoring.


Après quelques instants, l'écran montré ci-après s'affiche.

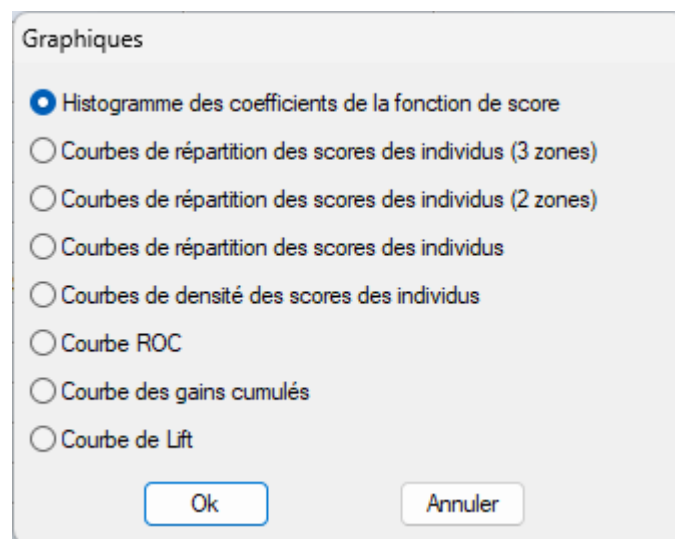
	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 10.2.0							
3								
4	DATE : 30/12/2024							
5	ORDINATEUR : LAPTOP-LEG8L077							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : CREDIT.SGD							
8								
9	RESULTATS DU SCORING							
10								
11	Sélection :							
12	Aucune							
13								
14	Nombre d'individus : 1000							
15								
16	Variables explicatives qualitatives :							
17	Compte chèque							
18	Durée							
19	Historique							
20	Objet							
21	Montant							


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


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



et l'icône 'Graphiques'  affiche la boîte de dialogue des options pour les graphiques :



La quatrième icône « Classement »  affiche un rapport contenant le classement des individus dans les groupes en fonction des taux ou seuils choisis dans les graphiques des courbes de répartition des scores des individus.

La cinquième icône « Simulation »  affiche un écran dans lequel, après avoir choisi un individu, il est possible de modifier ses modalités pour étudier l'évolution de son score et de son appartenance à une classe.

La sixième icône « Enregistrement »  permet l'enregistrement des principaux résultats issus de l'analyse.

La septième icône « Quitter »  permet de quitter l'analyse.

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme Explorateur, Général ou HTML et d'accéder à l'aide à l'interprétation par Intelligence Artificielle.

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

Le rapport affiche les informations suivantes :

- Liste des variables explicatives, facteur de classement et modalité attribuée au groupe des scores faibles
- Fonctions discriminantes issues de l'analyse des correspondances multiples effectuée sur l'ensemble des variables qualitatives (ACM) :
 - Coefficient de la fonction discriminante
 - Coefficient de régression
 - Erreur-type
 - t de Student
 - Probabilité

Les numéros des facteurs issus de l'ACM ayant des probabilités inférieures à 0,05 (confiance à 95% sur le coefficient de régression) sont indiqués.

- Tableau des inerties de l'analyse factorielle discriminante (AFD) :
 - Valeur propre
 - Corrélation canonique
 - Lambda de Wilks
 - Khi-2
 - Nombre de degrés de liberté
 - Niveau de signification

- Test de Pillai (trace, Fisher, degrés de liberté et probabilité)
- Fonctions discriminantes sélectionnées

Si l'option « Sélection des facteurs significatifs » n'a pas été sélectionnée, ce tableau n'est pas affiché. Si cette option a été sélectionnée, seuls les facteurs ayant des probabilités inférieures à 0,05 dans le précédent tableau sont présents.

- Fonctions discriminantes exprimées en fonction des modalités des variables explicatives qualitatives
 - Coefficient de la fonction discriminante
 - Coefficient de régression
 - Erreur-type
 - t de Student
 - Probabilité
- Classement issu de l'analyse factorielle discriminante (AFD)
 - Comptages et pourcentages : affiche les classements dans les groupes observés et prévus.
 - Statistiques : affiche le taux global d'erreur, la statistique Kappa, la sensibilité, la spécificité et le score F1.

Pour aider à la lecture de ces informations, notons :

VP : Vrais positifs
 VN : Vrais négatifs
 FP : Faux positifs (erreur de type I)
 FN : Faux négatifs (erreur de type II)

Par exemple, un « VP » est un « bon » client classé « bon », un « VN » est un « mauvais » client classé « mauvais », un « FP » est un « mauvais » client classé « bon » et un « FN » est un « bon » client classé « mauvais ».

Avec ces notations, nous pouvons élaborer le tableau suivant :

	Prévu Bon	Prévu Mauvais	Total	% correct	
Bon	<i>VP</i>	<i>FN</i>	<i>VP+FN</i>	$\frac{100 * VP}{(VP + FN)}$	Sensibilité
Mauvais	<i>FP</i>	<i>VN</i>	<i>FP+VN</i>	$\frac{100 * VN}{(VN + FP)}$	Spécificité
Total	<i>VP+FP</i>	<i>VN+FN</i>			
% correct	$\frac{100 * VP}{(VP + FP)}$	$\frac{100 * VN}{(VN + FN)}$		$\frac{100 * (VP + VN)}{(VP + VN + FP + FN)}$	
				% total correctement prévu	

Les statistiques affichées se définissent alors par les formules suivantes :

- Taux global d'erreur (%) : 100 - % total correctement prévu
- Kappa (%) : compare le classement obtenu à un classement obtenu par hasard.

$$\text{Kappa} = 100 * \frac{p(\text{correct}) - p(\text{hasard})}{1 - p(\text{hasard})} \text{ avec :}$$

$$p(\text{correct}) = \frac{(VP+VN)}{(VP+VN+FP+FN)}$$

$$p(\text{hasard}) = \frac{(VP - (\frac{VP+FP}{2})) + (VN - (\frac{VN+FN}{2}))}{(VP+VN+FP+FN)}$$

Le classement est parfait si Kappa = 100% et pas meilleur que celui obtenu par hasard si Kappa = 0%.

- Sensibilité (%) : $\frac{100*VP}{(VP+FN)}$

Plus elle est forte plus la proportion des vrais bons clients est grande parmi le total des bons clients. En d'autres termes, la sensibilité mesure la capacité de la fonction de classement à classer un client « positif » si ce client est « positif ».

- Spécificité (%) : $\frac{100*VN}{(VN+FP)}$

Plus elle est forte plus la proportion des vrais mauvais clients est grande parmi le total des mauvais clients. En d'autres termes, la spécificité mesure la capacité de la fonction de classement à classer un client « négatif » si ce client est « négatif ».

- Score F1 : $2 * \frac{Précision*Sensibilité}{Précision+ Sensibilité}$

C'est la moyenne harmonique de la précision et de la sensibilité. Elle est comprise entre 0 et 1. Elle mesure la qualité du classement lorsqu'il est attribué un même poids à la précision et à la sensibilité.

- Coefficients de la fonction de score

Ce tableau affiche les coefficients de la fonction discriminante ainsi que les coefficients de la fonction de score.

- Coordonnées et scores des individus

Ce tableau affiche les coordonnées et scores des individus ainsi que les groupes d'appartenance des individus (« Bon client » ou « Mauvais client »).

Les scores sont évalués sur une échelle de 0 à 1000 de la façon suivante : pour chaque variable qualitative, le plus petit coefficient est mis à la valeur 0. Ainsi le plus petit score (0) est obtenu par un individu qui pour chaque variable possède la modalité ayant un coefficient transformé égal à 0. Le score maximum (1000) correspond à la somme des coefficients transformés les plus grands dans chaque variable. Le score attribué à un individu s'obtient en additionnant les coefficients transformés associés aux modalités prises par l'individu.

- Seuil, score, VP, FN, FP, VN, sensibilité, spécificité

Ce tableau affiche pour chaque seuil, les informations suivantes : score, VP, FN, FP, VN, sensibilité et spécificité. Il indique également le niveau de qualité (aire sous la courbe). Ces informations sont également présentées sous forme graphique.

- Gains cumulés et valeurs de Lift

Ce tableau affiche pour chaque seuil les informations suivantes : score, pourcentage de tous les individus, sensibilité et valeur de Lift. Il indique également la valeur de l'indice de Gini. Ces informations sont également présentées sous forme graphique.

- Classement issu du scoring

Ce rapport n'est disponible que si les taux et les seuils ont été définis. Pour cela, il faut préalablement tracer les « Courbes de répartition des scores des individus ». Pour plus d'informations, voir l'option « Classement ».

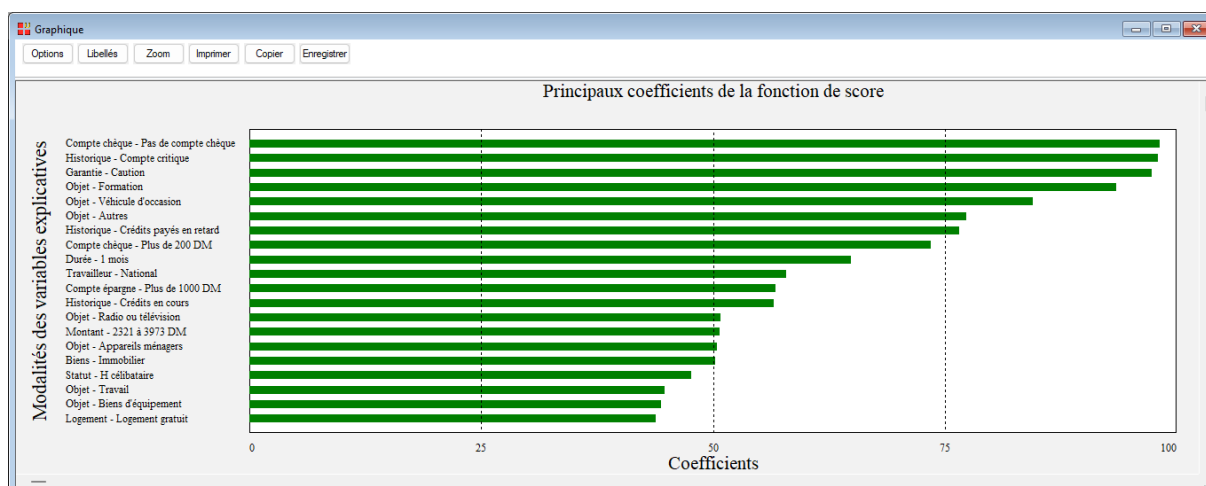
L'option Graphiques

Cette option permet d'obtenir divers graphiques présentés ci-après.

- Histogramme des coefficients de la fonction de score

Ce graphique affiche les plus grands coefficients de la fonction de score (20 au maximum) sous la forme d'un histogramme de ces coefficients triés par valeurs décroissantes.

Il permet ainsi de visualiser rapidement les modalités des variables qui ont les plus forts impacts sur les scores des individus.



- Courbes de répartition des scores des individus (3 zones)

Ce graphique affiche les deux courbes de répartition des scores des individus « bons » et « mauvais ».

Elle affiche également trois zones : la zone rouge, la zone orange et la zone verte. Les limites de ces zones sont définies par les taux rouge et vert ou par les seuils rouge et vert.

Les taux rouge et vert (taux d'erreur de classement) représentent respectivement le pourcentage de « bons » classés « mauvais » par la fonction de score et le pourcentage de « mauvais » classés « bons » par la fonction de score. Ils sont par défaut initialement définis tous les deux à environ 10%.

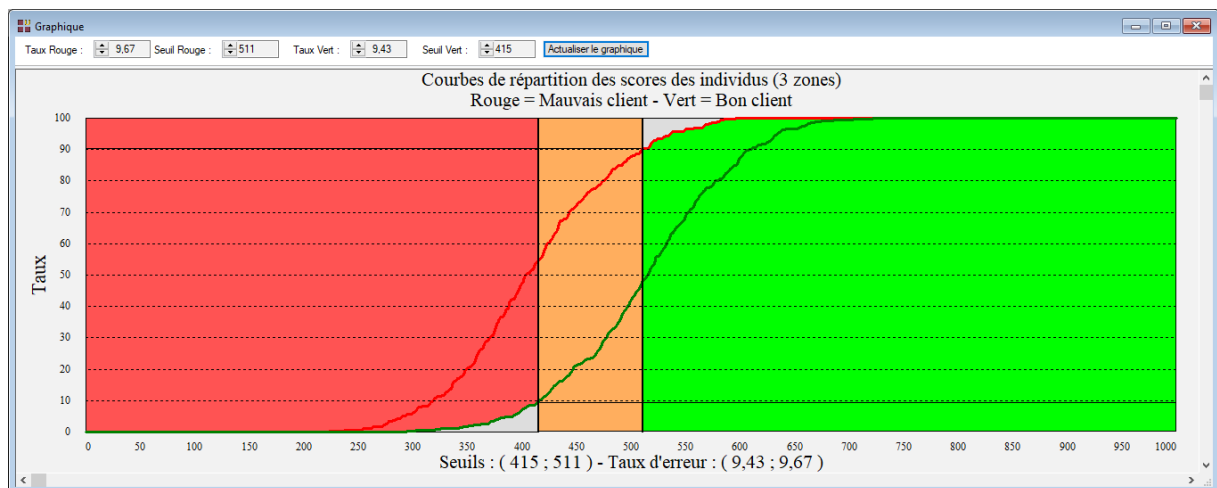
La zone rouge correspond aux individus des deux groupes déclarés « mauvais » par la fonction de score, à savoir le groupe des « mauvais » classés « mauvais » et le groupe des « bons » classés « mauvais ».

La zone orange correspond aux individus des deux groupes déclarés « non classés » par la fonction de score. C'est une zone d'indécision.

La zone verte correspond aux individus des deux groupes déclarés « bons » par la fonction de score, à savoir le groupe des « bons » classés « bons » et le groupe des « mauvais » classés « bons ».

Par exemple, le score 415 est tel que 9,43 % des individus du groupe des scores forts sont mal classés et le score 511 est tel que 9,67 % des individus du groupe des scores faibles sont mal classés.

Il est possible de modifier les valeurs des taux et des seuils en utilisant les boutons fléchés dans la barre d'outils au-dessus du graphique.

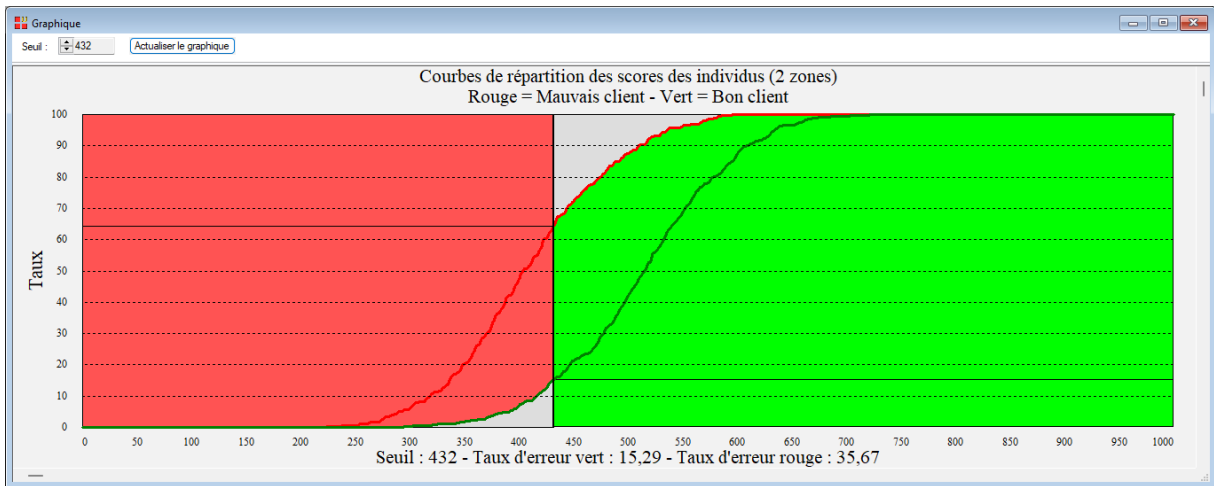


- Courbes de répartition des scores des individus (2 zones)

Ce graphique affiche les deux courbes de répartition des scores des individus « bons » et « mauvais ».

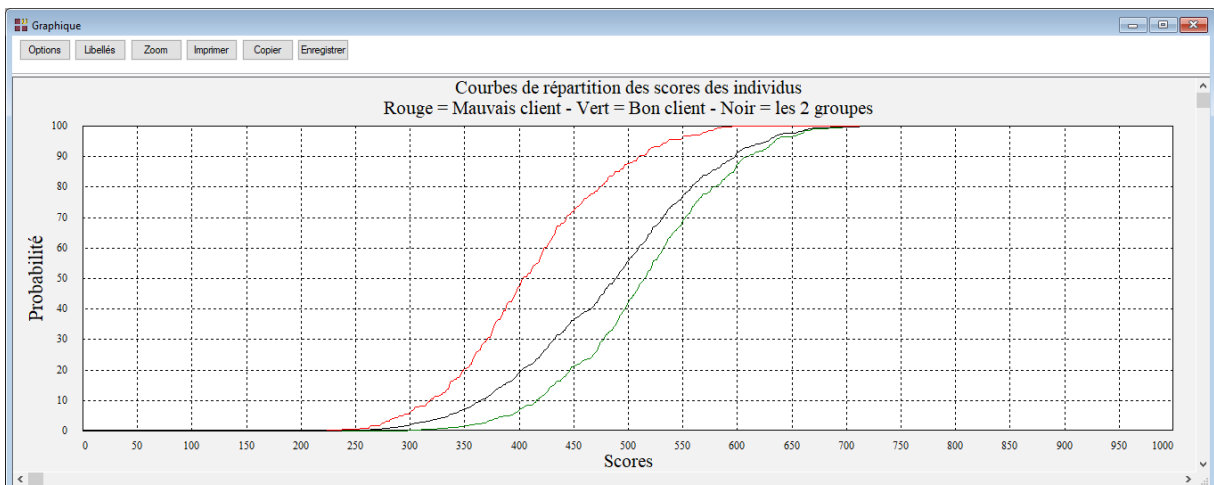
Elle affiche uniquement deux zones, rouge et verte, sans zone orange d'indécision. La limite séparant ces deux zones peut être modifiée en utilisant le bouton fléché dans la barre d'outils au-dessus du graphique.

Le seuil affiché par défaut (ici 432) correspond au score tel que 30% (300 « mauvais » sur 1.000 demandeurs) de tous les individus ont un score inférieur ou égal à ce seuil.



- Courbes de répartition des scores des individus

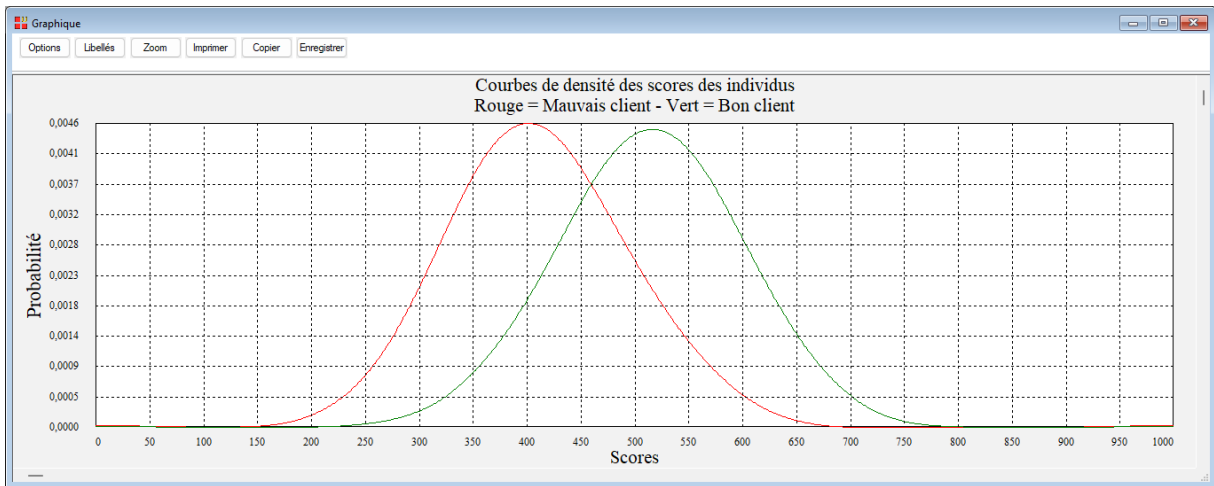
Ce graphique affiche les courbes de répartition des scores des individus pour les deux groupes « bons » et « mauvais » ainsi que la courbe de répartition globale.



- Courbes de densité des scores des individus

Ce graphique affiche les deux courbes de densité des scores des individus « bons » et « mauvais ».

Ces courbes sont obtenues par un lissage utilisant la méthode « cosinus » avec une largeur d'intervalle de 25 % (100 individus ou plus) ou 50% (moins de 100 individus).

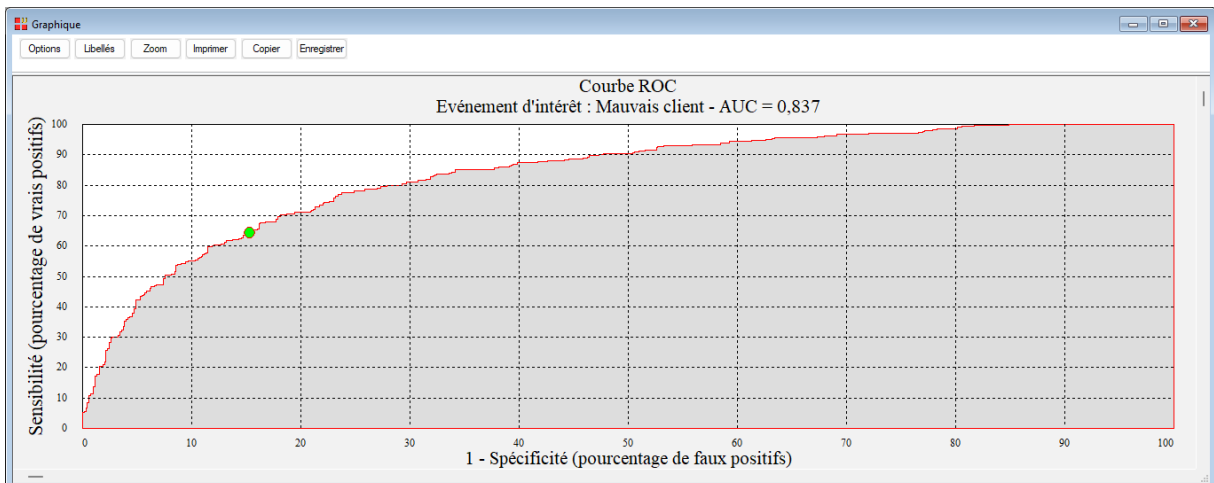


- Courbe ROC

La courbe ROC affiche la relation entre la sensibilité (proportion de vrais « bons » chez les « bons » ou vrais positifs) et $1 -$ la spécificité (proportion de faux « bons » chez les « mauvais » ou faux positifs).

Cette courbe répond à une problématique de risque : détecter le plus possible de positifs (avoir une fonction de score la plus sensible possible) tout en ayant le moins possible de fausses alertes. C'est par exemple le cas pour les assurances, les banques ou la médecine.

Plus cette courbe est proche de la partie supérieure du cadre du graphique, meilleure est la séparation entre les groupes réalisée par la fonction de score. Lorsque les densités des deux groupes sont identiques, la courbe ROC correspond à la diagonale.



En cliquant sur le point représenté par un cercle vert (score 432), il est possible de visualiser les valeurs correspondantes de la sensibilité et de $1 -$ la spécificité. Cela indique qu'au-dessus d'un score de 432, il y a environ 64 % de « vrais bons » (parmi les « bons ») et environ 15 % de « faux bons » (parmi les « mauvais »).

Il est également possible de cliquer en tout emplacement proche de la courbe pour visualiser ces informations pour différents seuils.

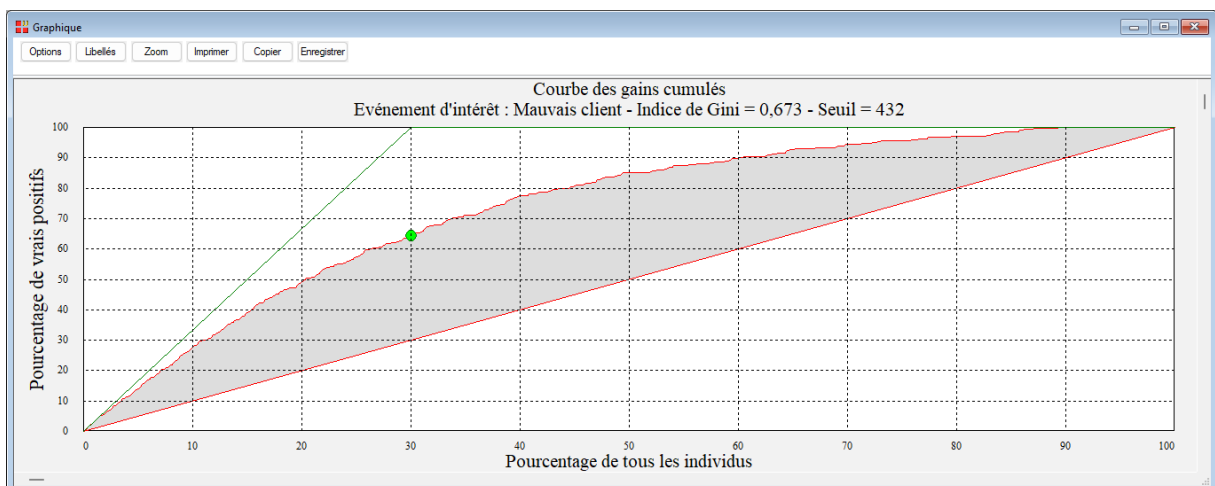
Le niveau de qualité, ici de 83,7 %, représente l'aire sous la courbe. Le meilleur niveau possible est de 100%. Un niveau de 50% correspond à un modèle pas meilleur que le hasard.

- Courbe des gains cumulés

La courbe des gains cumulés affiche la relation entre le pourcentage de vrais positifs et le pourcentage total des individus.

La courbe idéale est la ligne brisée (verte dans le graphique) qui correspond au cas où la distribution des « mauvais » est entièrement inférieure à la distribution des « bons ».

Cette courbe est souvent utilisée dans le milieu du marketing car elle répond à une question courante : avec un budget de campagne qui permet de cibler tel pourcentage de clients, quelle proportion d'acheteurs ? Une autre façon de poser le problème consiste à évaluer le pourcentage de clients à cibler, donc les moyens à prévoir, afin d'atteindre telle proportion des acheteurs habituels.



En cliquant sur le point représenté par un cercle vert (score 432), il est possible de visualiser les valeurs correspondantes des pourcentages de vrais positifs et du total des individus.

Il est indiqué que 30 % de tous les individus ont un score supérieur à 432 et que cela représente environ 64 % des vrais positifs.

Il est également possible de cliquer en tout emplacement proche de la courbe pour visualiser ces informations pour différents seuils.

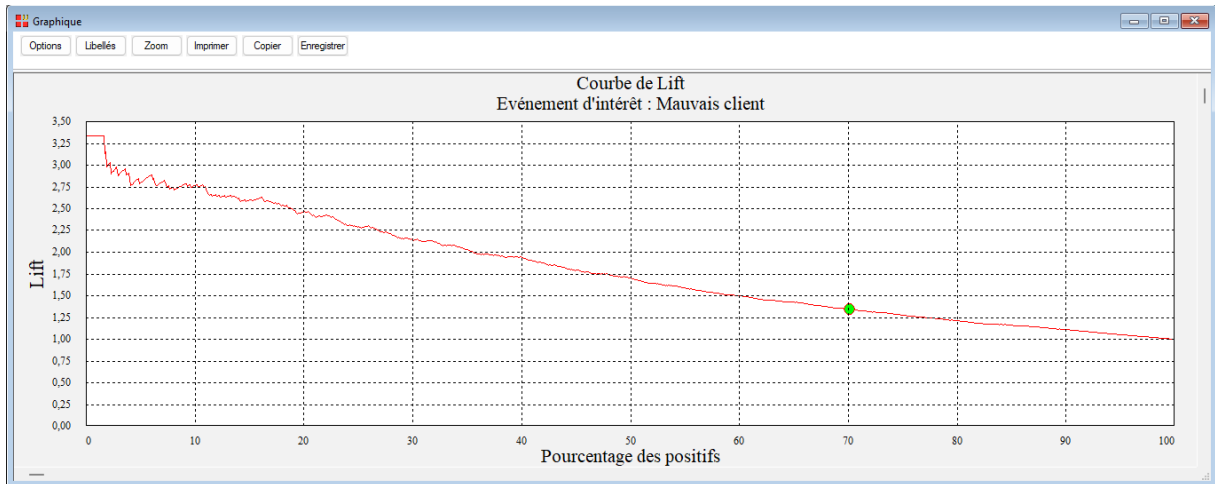
L'indice de Gini, ici de 0,673 représente la proportion grisée de l'aire entre la diagonale et la courbe idéale.

- Courbe de Lift

La courbe de Lift affiche la relation entre la valeur de Lift et le pourcentage des positifs :

$$\text{Pourcentage des positifs} = (VP+FP) / (VP+FP+VN+FN)$$

$$\text{Valeur de Lift} = \text{Sensibilité} / \text{Pourcentage des positifs}$$



En cliquant sur le point représenté par un cercle vert (score 4327, il est possible de visualiser les valeurs correspondantes de la valeur de Lift et du pourcentage des positifs.

Il est indiqué que pour 70 % des positifs, la valeur de Lift est égale à 1,34.

Il est également possible de cliquer en tout emplacement proche de la courbe pour visualiser ces informations pour différents seuils.

L'option Classement

Pour visualiser ce classement, il est possible de demander le rapport complet (première icône dans la barre d'outils) ou pour ne visualiser que ce classement de cliquer sur la quatrième icône de cette barre d'outils. Cette option est plus efficace lorsque vous souhaitez étudier plusieurs valeurs possibles pour les seuils.

Par ailleurs, ce rapport du classement n'est obtenu qu'après avoir tracé un des graphiques des courbes de répartition des scores des individus dans lesquels le ou les seuils des zones rouge et verte sont choisis.

Toute modification des seuils ou taux dans les graphiques des courbes de répartition provoque automatiquement la mise à jour de ce rapport et du rapport général.

Le rapport affiche les informations suivantes :

- les taux et seuils utilisés pour le classement
- les statistiques globales issues du classement (comptages et pourcentages)
- le classement pour chaque individu (score, origine, affectation, erreur)
 - En « bon », « mauvais » et « non classé » (si 3 zones)
 - En « bon » et « mauvais » (si 2 zones)

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	TAUX ET SEUILS UTILISES											
3												
4												
5												
6												
7												
8												
9												
10												
11	RESULTATS DU CLASSEMENT ISSU DU SCORING (COMPTAGES)											
12	TAUX ROUGE : 35,67 - SEUIL ROUGE : 432 - TAUX VERT : 15,29 - SEUIL VERT : 432											
13												
14	EN LIGNES, LES GROUPES OBSERVES											
15	EN COLONNES, LES GROUPES PREVUS											
16												
17												
18												
19												
20												
21												

L'option Simulation

L'option Simulation (cinquième icône de la barre d'outils) permet de visualiser l'évolution du score d'un individu lorsqu'une ou plusieurs de ses modalités sont modifiées.

Pour utiliser cette option, il suffit de sélectionner l'individu dans la liste déroulante de la barre d'outils.

Les modalités d'origine de l'individu pour les différentes variables s'éclairent et les groupes d'origine et d'affectation ainsi que le score actuel s'affichent.

Pour modifier une modalité pour une variable, il suffit de cliquer sur la ligne de cette modalité. Elle s'éclaire alors et les groupe d'origine et d'affectation ainsi que le score sont mis à jour.

A noter que les modifications effectuées lors de la simulation ne sont pas conservées et que le fichier des données n'est pas modifié.

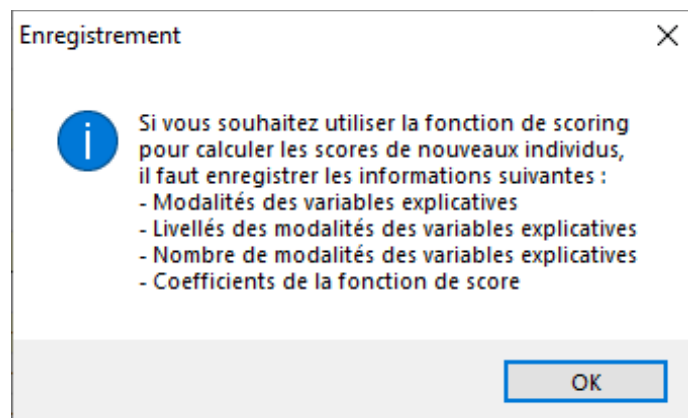
Rapports et Graphiques			
			Individu : INDA0001
1			
2	SIMULATION DU SCORE DE L'INDIVIDU : INDA0001		
3			
4	Origine	Affectation	Score
5	Bon client	Mauvais client	405,08
6			
7	Variable	Modalité	Coefficient
8	Graphique des coefficients de la fonction de score		
9	Compte chèque	Compte chèque - 0 DM	0,00
10	Compte chèque	Compte chèque - 1 à 200 DM	27,90
11	Compte chèque	Compte chèque - Pas de compte chèque	98,19
12	Compte chèque	Compte chèque - Plus de 200 DM	73,53
13			
14	Durée	Durée - 1 mois	64,84
15	Durée	Durée - 2 mois	38,61
16	Durée	Durée - 3 mois	31,85
17	Durée	Durée - 4 mois	0,00
18			
19	Historique	Historique - Compte critique	97,96
20	Historique	Historique - Crédits en cours	56,60
21	Historique	Historique - Crédits payés en retard	76,58

L'option Enregistrement

L'option Enregistrement permet d'enregistrer dans le fichier des résultats certaines informations issues de la procédure de scoring.

Ces informations sont automatiquement écrites dans une feuille du tableur.

Un message s'affiche précisant les informations à enregistrer si vous souhaitez utiliser la fonction de score pour calculer les scores d'une population de prévision.



Les informations qui peuvent être enregistrées sont :

- Modalités des variables explicatives
- Nombres de modalités des variables explicatives
- Libellés des modalités des variables explicatives
- Coefficients de la fonction de score
- Libellés des individus
- Groupes des individus
- Coordonnées des individus
- Scores des individus

- Seuils, Vrais positifs, Faux négatifs, Faux négatifs, Vrais négatifs, Spécificités, Sensibilités, Gains X et Y, Lift X et Y

Dans le cadre de notre exemple, enregistrons dans le fichier des résultats SCO_Crédit les informations nécessaires à l'étude d'une population de prévision.

Enregistrement des résultats (1/2)

Enregistrer

- * Modalités variables explicatives
- * Nb. modalités variables explicatives
- * Libellés modalités variables explicatives
- * Coefficients fonction de score
- Libellés des individus
- Groupes des individus
- Coordonnées des individus
- Scores des individus
- Seuils
- Vrais positifs

Noms attribués aux variables cibles

- ModVarExp
- NbModVarExp
- LibModVarExp
- CoefScore
- LibInd
- Grplnd
- CoordInd
- ScoreInd
- seuil
- vp

Ok Plus Tout Annuler

Population de prévision : Fichier Crédit2

Pour étudier une population de prévision, quittons la procédure en cours.

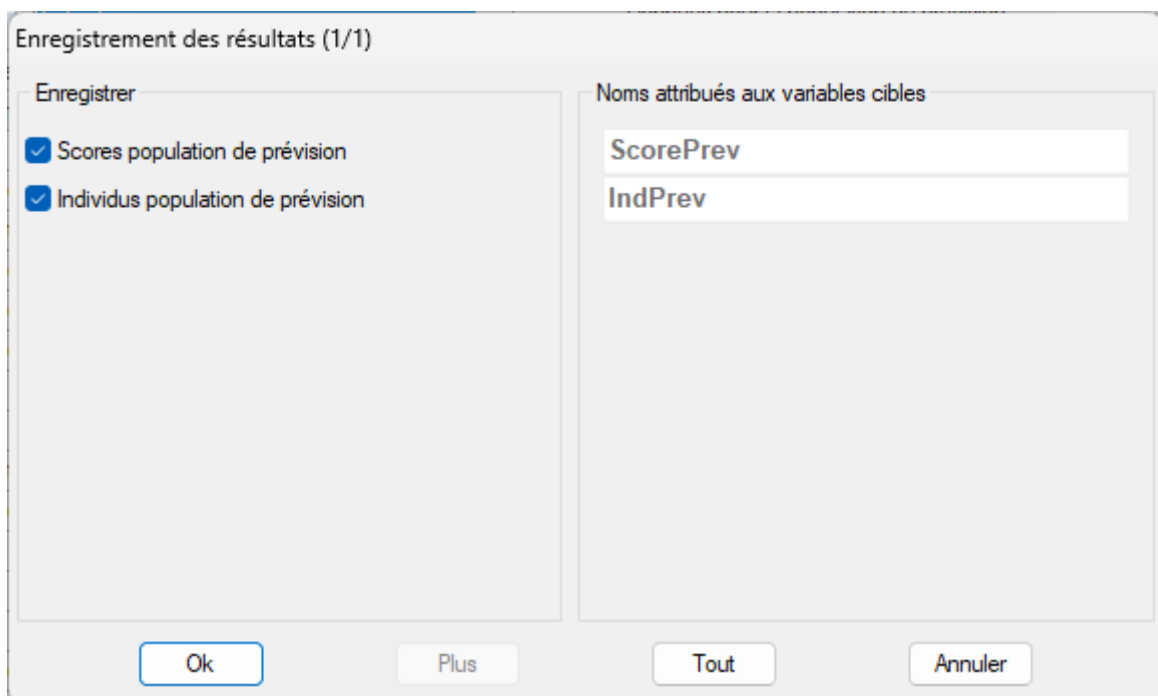
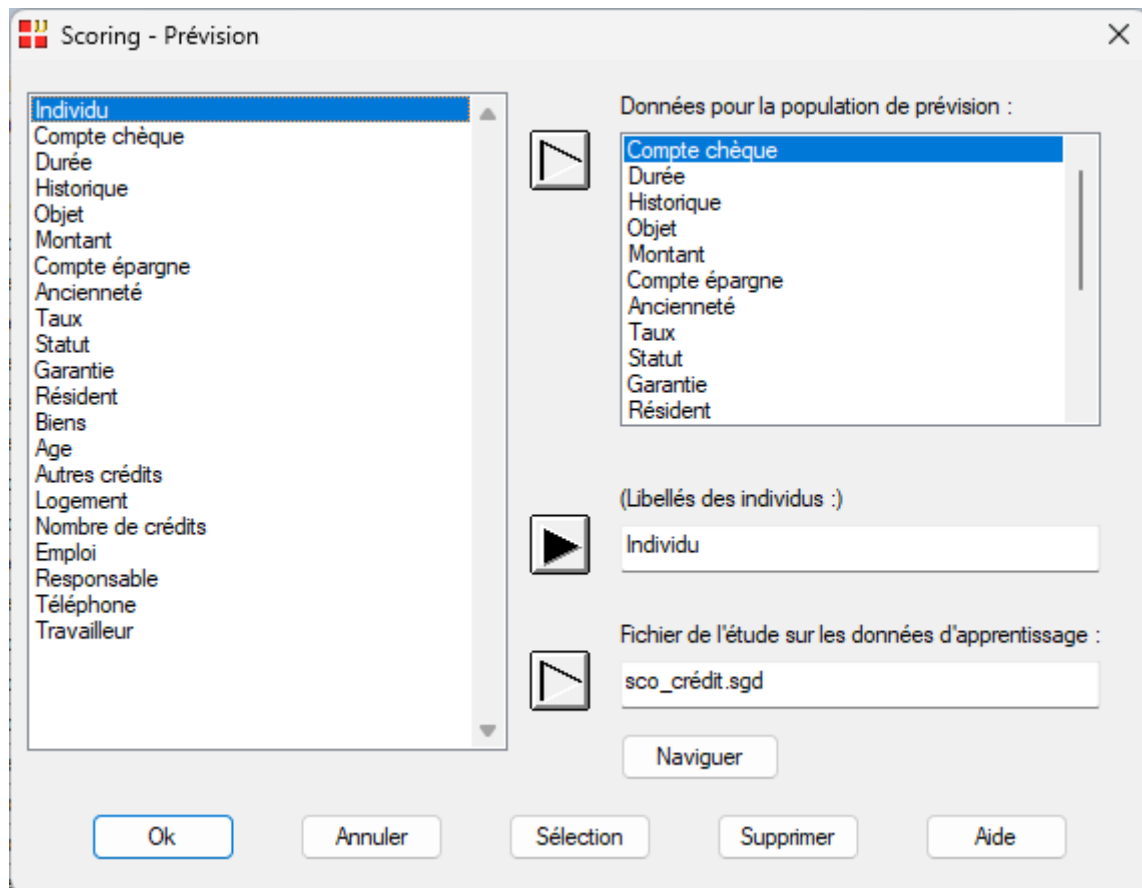
Ouvrons le fichier Crédit2 contenant la population de prévision puis affichons la boîte de dialogue d'entrée des données accessible par le ruban Expliquer – Scoring - Prévision.

Les individus de la population de prévision doivent être décrits par les mêmes variables que les individus de la population d'apprentissage et toutes les modalités utilisées par ces individus doivent avoir été présentes dans la population d'apprentissage.

Il faut préciser le nom du fichier de l'étude sur les données d'apprentissage contenant les résultats enregistrés issus de la population d'apprentissage, ici SCO_crédit.

Les scores calculés pour cette population de prévision peuvent être enregistrés dans le fichier des résultats.

Le calcul des scores des individus de la population de prévision est alors effectué et le rapport s'affiche.



	1	2	3	4	5	6	7	8
1								
2	SCORES DES INDIVIDUS DE LA POPULATION DE PREVISION							
3								
4		Score						
5	i1	490						
6	i2	633						
7	i3	262						
8	i4	521						
9	i5	393						
10	i6	533						
11	i7	544						
12	i8	657						
13	i9	245						
14	i10	411						
15	i11	594						
16	i12	523						
17	i13	478						
18	i14	434						
19	i15	544						
20	i16	438						
21	i17	474						

Exemple 2 : Fichier Ventes

Une enquête a été effectuée concernant 1596 clientes d'une société de vente par correspondance. Les clientes ont été interrogées sur les 22 thèmes suivants :

Thème

Libellé

- | | |
|--|--------------------|
| 1. Commander d'autres produits | Autres commandes |
| 2. Conseiller les produits à des amies | Conseils Amies |
| 3. Gestion des retours/échanges | Echanges |
| 4. Exactitude des factures | Factures |
| 5. Conformité du contenu du colis | Colis |
| 6. Rapidité de la livraison | Rapidité livraison |
| 7. Qualité de la livraison | Qualité livraison |
| 8. Diversité des modes de règlement | Règlement |
| 9. Flexibilité en cas d'annulation d'une commande | Annulation |
| 10. Facilité à passer une commande | Commande |
| 11. Montant des frais de port et d'emballage | Port |
| 12. Conformité entre photo et réalité des produits | Conformité |
| 13. Attrait, qualité et diversité des cadeaux | Cadeaux |
| 14. Possibilité d'essayer les produits | Echantillons |
| 15. Prix des produits | Prix |
| 16. Diversité et renouvellement des produits | Choix |
| 17. Qualité et efficacité des produits | Qualité produit |
| 18. Attrait et caractère innovant des produits | Attrait produit |
| 19. Attrait et clarté des offres promotionnelles | Promotions |
| 20. Relation téléphonique pour renseignements | Conseil téléphone |
| 21. Relation téléphonique pour commande | Commande téléphone |
| 22. Prise de contact en cas de problèmes | Prise Contact |

Chaque thème est codé de 1 à 5 :

- 1 = Très satisfaite
- 2 = Satisfaite
- 3 = Indifférente
- 4 = Mécontente
- 5 = Très mécontente

A noter que plusieurs thèmes ne présentent pas tous les codes de 1 à 5.

L'âge des clientes est codé de la façon suivante : plus de 50 ans, 30 à 50 ans, moins de 30 ans

La satisfaction des clientes est codée de la façon suivante :

- 1 = Très satisfaite
- 2 = Autres (satisfaite, indifférente, mécontente, très mécontente)

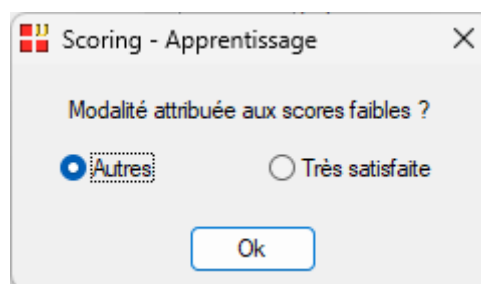
Cliquons sur l'icône SCO ou choisissons Scoring - Apprentissage dans le ruban Expliquer.

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

Sélectionnons les variables « Autres commandes » à « Age » comme variables explicatives qualitatives, la variable « Satisfaction » comme facteur de classement et la variable « Individu » comme variable contenant les libellés des individus.

Cliquons sur le bouton OK.

UNIWIN nous demande alors de préciser la modalité du facteur de classement qui doit être attribuée au groupe des scores faibles.



Indiquons que ce groupe est caractérisé par la modalité « Autres » du facteur de classement.

UNIWIN affiche alors les rapports et les graphiques de notre analyse.

Voici ci-dessous deux exemples des résultats affichés par l'explorateur.

Scoring - Apprentissage

Individu

Satisfaction

Autres Commandes

Conseils Amies

Echanges

Factures

Colis

Rapidité livraison

Qualité livraison

Règlement

Annulation

Commande

Port

Conformité

Cadeaux

Echantillons

Prix

Choix

Qualité produit

Attrait produit

Promotions

Conseil téléphone

Commande téléphone

Prise Contact

Age

Facteur de classement :

Satisfaction

Variables explicatives qualitatives :

Autres Commandes

Conseils Amies

Echanges

Factures

Colis

Rapidité livraison

Qualité livraison

Règlement

Annulation

Commande

Port

(Libellés des variables explicatives :)

(Libellés des individus :)

Individu

Nombre maximum de facteurs ACM : 100

Sélection des facteurs significatifs

Ok Annuler Sélection Supprimer Aide

Rapports et Graphiques

Individu : IND0001

Rapport Scoring

Variables utilisées

Fct. discr. (Issue de l'ACM)

Tableau des inerties AFD

Test de Wilco

Fct. discr. (sur modalités)

Classement issu de l'AFD

Fonction de score

Coordonnées et scores des individus

VP, FN, FP, VN, Sensibilité, Spécificité

Gains cumulés et valeurs de Lift

Classement issu du Scoring

	1	2	3	4	5	6	7	8
1								
2	FONCTION DISCRIMINANTE EXPRIMEE EN FONCTION DES MODALITES DES VARIABLES EXPLICATIVES							
3								
4								
5		Coef. fct. discr.	Coef. régression	Erreur-type	Student	Probabilité		
6	Constante	0,00000	0,00000	0,00000	0,00000	1,00000		
7								
8	Autres Commandes - 1	2,75706	3,49811	1,14130	3,06501	0,00218		
9	Autres Commandes - 2	-5,49724	-6,97479	2,23528	3,12033	0,00181		
10	Autres Commandes - 3	-2,63349	-3,34132	4,21705	0,79234	0,42816		
11	Autres Commandes - 4	-4,27183	-5,42002	13,46758	0,40245	0,68735		
12	Autres Commandes - 5	-9,99820	-12,68552	56,14956	0,22592	0,82126		
13								
14	Conseils Amies - 1	3,66404	4,64866	1,67950	2,76801	0,00564		
15	Conseils Amies - 2	-2,95429	-3,74834	1,71862	2,18102	0,02918		
16	Conseils Amies - 3	-2,77999	-3,52720	2,90672	1,21346	0,22495		
17	Conseils Amies - 4	-2,74434	-3,48197	8,33115	0,41795	0,67598		
18	Conseils Amies - 5	9,04937	11,48167	26,22606	0,43780	0,66153		
19								
20	Echanges - 1	0,34122	0,43294	3,79759	0,11400	0,90923		
21	Echanges - 2	1,31749	1,67161	2,57082	0,65022	0,51555		

Rapport Explorateur

Rapports et Graphiques

Individu : IND0001

Rapport Scoring

- Variables utilisées
- Fct. discr. (issue de l'ACM)
- Tableau des inerties AFD
- Test de Pillai
- Fct. discr. (sur modalités)
- Classement issu de l'AFD
- Fonction de score**
- Coordonnées et scores des individus
- VP, FN, FP, VN, Sensibilité, Spécificité
- Gains cumulés et valeurs de Lift
- Classement issu du Scoring

	1	2	3	4	5	6	7	8
1								
2	COEFFICIENTS DE LA FONCTION DE SCORE							
3								
4	Modalité attribuée au groupe des scores faibles : Autres							
5								
6								
7		Coef. fct. discr.	Coef. fct. score					
8	Autres Commandes - 1	2,75706	54,56265					
9	Autres Commandes - 2	-5,49724	19,25356					
10	Autres Commandes - 3	-2,63349	31,50371					
11	Autres Commandes - 4	-4,27183	24,49545					
12	Autres Commandes - 5	-9,99820	0,00000					
13								
14	Conseils Amies - 1	3,66404	28,31094					
15	Conseils Amies - 2	-2,95429	0,00000					
16	Conseils Amies - 3	-2,77999	0,74558					
17	Conseils Amies - 4	-2,74434	0,89809					
18	Conseils Amies - 5	9,04937	51,34756					
19								
20	Echanges - 1	0,34122	11,75683					
21	Echanges - 2	1,31749	15,93297					

Rapport Explorateur

Le graphique des coefficients de la fonction de score nous indique que les principales variables définissant le score d'une cliente sont la qualité des produits, le choix des produits, la relation téléphonique lors de la commande, la flexibilité d'annulation d'une commande et le montant des frais de port.

