

ROYAUME DU MAROC
*_*_*_*_*
PREMIER MINISTRE
*_*_*_*_*_*_*_*_*
HAUT COMMISSARIAT AU PLAN
_
INSTITUT NATIONAL
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE

INSEA



Projet de Fin d'Etudes

La Notation Interne selon Les Nouvelles Recommandations de Bâle II (Crédit Scoring).

Préparé par: Mr.Mohamed Salah CHIADMI
Mr.Hassan BOUNA

Sous la direction de : Mr.A.Chaoubi (INSEA)
Mr.Chemaou Mokhtar (Crédit agricole du Maroc)

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

**Option : Actuariat finance
Statistique**

Devant le jury composé de :

- Mr. A.CHAOUBI (INSEA)
- Mr. Mohamed TAAMOUTI (INSEA)
- Mr. Mokhtar CHEMAOU (Crédit agricole du Maroc)

Juin 2007

Dédicace

A mes chers parents symboles de sacrifice et d'amour

A mes chères frères : Houssine, Jaouad, Jamal et Saad

A toute ma famille

A mon binôme Salah CHIADMI

A tous mes amis

A toutes les personnes qui ont contribué de près ou de loin à la réussite de ce modeste travail.

A vous tous merci.

Hassan BOUNA

Dédicace

Je dédie ce petit travail

A mon binôme Hassan BOUNA,

A tous mes proches,

A tous ceux qui crient haut et fort contre l'injustice et la dictature.

Mohammed Salah

Remerciement

Au terme de ce travail, nous tenons à exprimer notre profonde gratitude à Monsieur CHAOUBI, notre encadrant à l'INSEA pour sa patience, ses précieux conseils qui nous ont été d'une très grande utilité et pour l'intérêt qu'il a accordé à notre travail durant toute la période de notre projet.

Nous remercions également Monsieur Mokhtar CHEMAOU, notre encadrant au Crédit agricole du Maroc pour nous avoir accueilli parmi son équipe pendant deux mois et demi de stage et pour le grand soutien et les grands efforts qu'il n'a pas économisés à nous accorder.

Nous remercions également M. Professeur à l'INSEA, pour avoir été parmi les membres de jury.

Enfin, nous adressons nos vifs remerciements au personnel du Département audit et qualité du risque au Crédit agricole du Maroc, et à tous ceux qui ont contribué de près ou de loin à la réussite de ce travail.

Merci infiniment !

Résumé

En vue d'une maîtrise optimale des risques encourus par les institutions financières, notamment le risque de crédit, l'amélioration du système de surveillance prudentielle s'avère fort nécessaire à cause des profondes mutations que connaît le paysage financier actuellement. C'est dans cette optique que s'inscrit notre projet de fin d'études à travers lequel nous avons traité la problématique de la minimisation des risques d'impayés afin d'optimiser les fonds propres de l'organisme en question (le CAM). C'est ainsi que nous avons engagé une étude basée sur l'approche dite de notation interne conforme aux normes bancaires internationales dictées par le comité Bâle II. Ainsi, à travers cette approche, nous avons abouti à élaborer deux modèles, le premier basé sur la méthode DISQUAL, et le deuxième sur la Régression Logistique, qui nous ont permis à la fois de calculer les scores des modalités, les probabilités de défaut associées et d'identifier les classes de risques. Nous avons constaté à la fin de notre application que le premier modèle est plus pertinent que le deuxième après avoir effectué plusieurs tests. Enfin, nous avons achevé notre étude par une automatisation du modèle DISQUAL à travers une application à l'aide du logiciel VB-Excel.

Mots clés : Risque de crédit, Bâle II, ACM, AFD, CHA, Régression logistique, Disqual, classe de risque, score, PDM, Fonds propres.

Table des Matières

Dédicace	1
Remerciement	3
Résumé	4
Liste des Abréviations	7
Liste des Tableaux	8
Liste des Figures	9
Introduction	10
Problématique	12

PREMIERE PARTIE : Présentation Générale

Chapitre 1 : Présentation de l'organisme d'accueil (Le Crédit Agricole du Maroc)	
I. Missions	14
II. Historique	14
III. Organisation	15
Chapitre 2 : Présentation de Bâle I et Bâle II	16
I. Les principaux risques bancaires	16
I.1. Risque de crédit	16
I.2. Risque de marché	16
I.3. Risque opérationnel	16
II. Présentation du Bâle I	17
III. Présentation des nouvelles recommandations de Bâle II	17
III.1. Les exigences minimales en fonds propres	18
III.2. Processus de surveillance prudentielle	18
III.3. Discipline du marché	19
Chapitre 3 : Définition du défaut et processus d'octroi et de recouvrement	20
I. Définition du défaut	20
II. Processus d'octroi et de recouvrement	20

DEUXIEME PARTIE : Présentation de la base de données

I. Elaboration de la base de données	22
I.1. Définition du portefeuille d'étude	22
I.2. Les étapes de la construction de la base de données	22
I.3. Les variables retenues pour l'étude	25
II. Etude descriptive	25
II.1. Statistique univariées	25
II.2. Les modalités des critères retenus	27
II.3. Répartition de clients	28
II.4. Tests des hypothèses de normalité et d'indépendance	30
II.4.1. Test de la normalité des variables retenues	30
II.4.2. Test de l'indépendance	30

TROISIEME PARTIE : Approche Théorique

Chapitre 1 : Méthode Disqual	33
Chapitre 2 : Régression Logistique	39
Chapitre 3 : Les techniques de validation du modèle	44

QUATRIEME PARTIE : Application

Chapitre 1 : Elaboration du modèle de Scoring par la méthode Disqual	49
I. Application de l'ACM	49
II. Application de l'AFD	50
III. Les scores associés à chaque modalité	52
IV. Validation du modèle sur l'échantillon d'apprentissage	54
IV.1. Taux de bons classements	54
IV.2. Boîtes à moustaches	55
IV.3. La courbe de ROC	55
V. Classe de risque et seuil d'affectation	56
V.1. Classe de risque	56
V.2. Le seuil d'affectation	57
Chapitre 2 : Elaboration du modèle de Scoring par la Régression logistique	59
I. Présentation de la base de données	59
II. Construction du modèle logistique	60
II.1. Sélection des variables du modèle	60
II.2. Significativité du modèle	60
II.2.1. Test de Wald	60
II.2.2. Test du rapport de vraisemblance	61
II.2.3. Autres tests de validation générale	61
III. validation du modèle sur l'échantillon de construction	61
III.1. Taux de bons classements	61
III.2. La courbe de ROC	62
III.3. Les boîtes à moustaches	63
IV. Les classes d'affectations	63
V. Simulation de seuil d'affectation	64
VI. Comparaison entre le modèle logistique et Disqual	65
VI.1. Taux de bons classements	65
VI.2. La courbe de ROC	65

CINQUIEME PARTIE : Informatisation du modèle de Scoring (DISQUAL) 68

Conclusion	72
ANNEXE I	73
ANNEXE II	75
Bibliographie	81

Liste des Abréviations

ACM : analyse de correspondance multiple

AFC : analyse factorielle de correspondance

AFD : analyse factorielle discriminante

CAM : crédit agricole du Maroc

CHA : classification hiérarchique automatique

PDM : probabilité de défaut moyenne

PFE : Projet de Fin d'étude

Liste des Tableaux

- **Tableau1** : caractéristiques des clients en défauts
- **Tableau2** : caractéristiques des clients bons
- **Tableau3** : répartition des deux types de clients par modalité
- **Tableau4** : significativité de la normalité des variables
- **Tableau5** : significativité de l'indépendance des variables
- **Tableau6** : les valeurs propres associés aux axes factoriels
- **Tableau7** : les axes factoriels les plus discriminants
- **Tableau8** : coefficients des axes les plus discriminants
- **Tableau9** : significativité du modèle
- **Tableau10** : les scores des modalités
- **Tableau11** : caractéristiques des deux catégories de clients
- **Tableau12** : Matrice de confusion (échantillon d'apprentissage)
- **Tableau13** : classe de risque et la probabilité de défaut associée
- **Tableau14** : les classes de risques homogènes et PD
- **Tableau15** : les stratégies d'affectations des clients
- **Tableau16** : la règle de décision
- **Tableau17** : les modalités supprimées (références)
- **Tableau18** : les coefficients des variables retenues
- **Tableau19** : la statistique de Wald pour les variables retenues
- **Tableau20** : taux de bons classements des deux groupes de clients
- **Tableau21** : le taux de bons classements pour les méthodes de discriminations
- **Tableau22** : indice de Gini pour les deux méthodes de discriminations
- **Tableau23** : classe de risque par la méthode logistique
- **Tableau24** : classes de risques et leurs PD (Logistique)
- **Tableau25** : stratégies d'affectation des clients
- **Tableau26** : règle de décision
- **Tableau27** : la forme de la base de donnée finale de travail pour la régression logistique
- **Tableau28** : coordonnées des modalités sur les axes les plus discriminants
- **Tableau29** : coefficients (scores) des axes factoriels
- **Tableau30** : règle de décision (d'affectation)
- **Tableau31** : corrélation canonique de l'analyse discriminante pour les deux groupes
- **Tableau 32** : quelques critères de comparaison entre les méthodes

Liste des Graphiques

- **Graphique1** : graphe de valeurs propres associés aux axes factoriels
- **Graphique2** : graphe de dispersion des clients selon F10 et F7
- **Graphique 3** : Boîte à moustaches pour les deux groupes de clients (Disqual)
- **Graphique 4** : courbe de ROC (méthode Disqual)
- **Graphique 5** : courbe de ROC pour les deux groupes de clients (Logistique)
- **Graphique 6** : Boîte à moustaches pour les deux groupes de clients (Logistique)
- **Graphique 7** : comparaison entre les deux méthodes de discrimination par la courbe de ROC
- **Graphique 8** : page d'accueil de l'interface d'automatisation du modèle de notation
- **Graphique 9** : page d'introduction des informations préliminaires du client
- **Graphique 10** : page d'introduction des informations financières du client
- **Graphique 11** : la page de notation du client
- **Graphique 12** : la page représentative de la règle de décision
- **Graphique 13** : Architecture de la réforme Bâle II
- **Graphique 14** : dispersion des clients selon F10, F7, F13 et F21
- **Graphique 15** : Organigramme de l'organisme d'accueil (CAM)
- **Graphique 16** : Courbes de moyenne par groupe de clients
- **Graphique 17** : Les moyennes des variables (II) par type de clients

Introduction

Afin d'améliorer et promouvoir la qualité de l'actuel système de surveillance prudentielle, les autorités en charge de régulation et de réforme bancaire internationale ont engagé un nouveau processus qui couvre explicitement la gestion de trois types de risque: le risque de crédit, le risque de marché et le risque opérationnel et ce, dans le cadre du nouvel accord piloté par la banque des règlements internationaux Bâle II. Il s'agit alors de conforter la robustesse et la stabilité du système bancaire international en obligeant les banques à se prémunir contre une prise de risques trop lourde. C'est dans ce sens que le comité Bâlois recommande aux institutions financières de mettre en place de nouveaux systèmes d'évaluation et de mesure de risques qui devraient respecter les 3 piliers suivants dictés par le comité:

1. Exigences minimales en fonds propres pour couvrir les actifs pondérés en fonction du risque.
2. Contrôle des procédures et de la méthode interne d'affectation des fonds propres (processus de surveillance prudentielle).
3. Les établissements sont soumis à la discipline de marché et tenus de publier les informations complètes sur la nature, le volume et les méthodes de gestion de leurs risques aussi que sur l'adéquation de leurs fonds propres.

En effet, depuis 1988, un ratio international de solvabilité, le ratio Cooke, est appliqué par les banques. Il définit les exigences en fonds propres qu'elles doivent respecter en fonction des risques encourus. Bien que ce ratio présente actuellement certaines faiblesses du fait qu'il n'intègre pas le fonctionnement interne de la banque, le comité Bâlois a proposé en 1999 la mise en place d'un nouvel ratio, dit ratio de Mc Donough qui repose sur le 1 pilier cité ci-dessus.

Le nouvel accord encourage la mise en place des approches dites de notations internes pour la gestion du risque de crédit. De ce fait, le Crédit Agricole du Maroc, à la réponse des exigences dictées par Bank Al Maghrib et dans le cadre de la modernisation de son système d'information, envisage d'élaborer un nouveau processus de maîtrise de risques à travers la construction d'un modèle de notation interne conforme aux normes bancaires internationales. En effet, La mesure de risques repose sur plusieurs méthodes statistiques telles l'analyse linéaire discriminante de Fisher, la méthode de Disqual, la régression logistique ou modèle logit...(voire le tableau 32)

Dans le présent sujet, nous allons nous contenter de traiter notre problématique en procédant par les deux méthodes: la méthode de DISQUAL et la régression logistique. Ces deux méthodes permettraient à la fois de déterminer les classes de risques et les probabilités de défauts correspondantes.

Ainsi, notre étude va se reposer sur 5 axes principaux qui aborderont l'approche en mettant l'accent sur les points suivants:

- ☞ Cadre global et théorique.
- ☞ Identification de notre problématique et précision des risques des clients tombés en impayé.
- ☞ Séparation des clients « BONS » et des « DEFAULTS ».

- ☞ Élaboration d'un modèle de scoring qui permettra de noter les clients, identifier les classes de risques et calculer les probabilités de défaut associées.
- ☞ Attribution de décisions appropriées à chaque classe de risque.

Notre projet sera constitué de cinq principales parties. La première abordera la présentation du Crédit Agricole du Maroc et son processus d'octroi et de recouvrement adopté. Puis dans la deuxième partie, nous nous intéresserons aux étapes suivies pour composer notre base de données d'étude. Il sera question dans la troisième partie d'aborder le cadre théorique. Dans la quatrième partie, nous effectuerons une construction de la classification du modèle, une fois les modèles élaborés, nous effectuerons un ensemble de tests de validation et de performance, Ensuite, une comparaison entre les deux modèles issus de la méthode Disqual et la Régression Logistique sera traitée à la fin de cette partie. Enfin, dans la cinquième partie, Nous terminerons par une application informatique, à l'aide du logiciel Visual Basic, automatisant le modèle de scoring : Disqual proposé par l'organisme d'accueil. Cette application facilitera à la société la notation automatique des clients tombés en impayé selon leurs caractéristiques et à identifier leurs groupes de risque et les probabilités de défaut associées.

Problématique

Dans le cadre de la nouvelle réforme bancaire recommandée par le comité Bâlois (BâleII), les institutions financières ont entamé une nouvelle phase au niveau de la réorganisation de leur système d'information qui va avec l'amélioration de leur processus prudentiel pour assurer une gestion optimale des risques encourus notamment le risque de crédit, à travers la détermination des classes de risque et la probabilité de défaut correspondante, et par la suite l'optimisation des fonds propres. C'est dans ce sens que le Crédit Agricole du Maroc, à la réponse des exigences de BAM, envisage d'adopter un système de notation interne de sa clientèle.

Ainsi, notre présent sujet va aborder l'approche dite de notation interne dictée par BAM à travers l'élaboration d'un modèle fondé sur des méthodes statistiques à savoir la méthode DISQUAL et la régression logistique qui nous permettront de discriminer les clients selon l'identification des classes de risques et le calcul des probabilités de défaut associées.



Projet fin d'étude
2006-2007

Présentation Générale

Chapitre1 : Présentation de l'organisme d'accueil (CAM)

I. MISSIONS

Acteur majeur dans le développement de l'agriculture et la modernisation du monde rural, le Crédit Agricole joue, depuis sa création en 1961, un rôle essentiel dans le financement de l'agriculture et dans la consolidation du secteur dans son ensemble. Le Crédit agricole pérennise ce rôle de promotion du monde rural à travers son développement de la banque Universelle.

Par ailleurs, CAP 2008, le nouveau plan d'entreprise du Crédit Agricole du Maroc, se propose d'afficher ses ambitions stratégiques pour lui permettre d'envisager l'avenir avec sérénité et confiance, sa mission est triple :

- Définir une stratégie de déploiement de l'activité de la banque afin de pérenniser et de sécuriser sa croissance.
- Modifier en profondeur la culture de l'entreprise pour ancrer dans le quotidien le réflexe de l'optimisation du rapport rendement / risque.
- Mettre en place des outils de gestion qui souscriront la banque au standard des normes de la place.

II. HISTORIQUE

Création de la CNCA (1961-1967)

- en fonction de la Caisse Nationale de Crédit Agricole.
- Implantation des caisses locales sous formes de succursales.

Début de l'activité bancaire (1970-1987)

- Lancement de la collecte des dépôts et des activités bancaires.
- Financement de l'agro-industrie en 1979.
- Réorientation de la CNCA et financement de nouveaux secteurs en 1987 dont la pêche côtière, l'artisanat, le tourisme, le commerce et les services...

Nouvelles dispositions pour la CNCA (1988-1996)

- Impôt sur les sociétés.
- Dispositions de la nouvelle loi bancaire.
- Mise en place de la salle des marchés en 1996.

Nouveau positionnement stratégique de la CNCA (1997-2001)

(Banque rurale de proximité avec une nouvelle identité visuelle) :

- Mise en place des directions de réseau décentralisées.
- Plan d'entreprise « Oufok » 2003 lancé en 1999 : une nouvelle stratégie visant principalement à mettre à niveau l'institution et à engager les actions de redressement de sa situation financière et ce, dans le cadre d'un positionnement rénové en tant que banque rurale généraliste de proximité, agissant en partenaire actif de toute la filière agricole et de l'ensemble du monde rural.

- Adoption de la nouvelle réforme institutionnelle du Crédit Agricole par le conseil des ministres en avril 1999.
- Signature de la convention État-CNCA relative au traitement du surendettement des agriculteurs en 2001.

Extension du réseau du Groupe Crédit Agricole (2003-2004)

- Acquisition du Réseau BMAO.
- Acquisition du Réseau BNDE.

Changement de statut (2004)

- La CNCA devient Société Anonyme à Directoire et à Conseil de Surveillance, dénommée « Crédit Agricole du Maroc » régie par la loi relative aux sociétés anonymes ainsi que par la loi relative à l'exercice de l'activité des établissements de crédit et de leur contrôle.
- Achat du siège de la BNDE.
- Fusion/absorption BMAO.

III. ORGANISATION

Une nouvelle organisation, en phase avec les ambitions de la banque et sa stratégie de développement, a été élaborée en 2004 (voire figure 15). L'organigramme mis en place, s'inscrit dans la logique du projet d'entreprise « CAP 2008 » et répond à ses principes directeurs qui sont :

- Une Organisation par marché.
- Un Professionnalisation des métiers de la banque.
- Un renforcement du management et de la culture de la performance.

Notre projet de fin d'études a été effectué au sein du Département audit et Qualité du risque.

Chapitre2 : Présentation de Bâle I et Bâle II

La stabilité et la solvabilité du système bancaire sont une condition sine qua non pour le bon fonctionnement du système financier. Afin d'adapter et d'augmenter la souplesse de l'actuel système de surveillance, les autorités en charge de la régulation bancaire internationale ont engagé un processus de réforme du calcul des fonds propres nécessaires à la couverture des risques et ce, dans le cadre du comité de Bâle II, mais avant de présenter les nouvelles recommandations concernant la gestion des risques, on va présenter le déroulement historique des recommandations du comité Bâle dictées pour les institutions financières en vue de cerner les différents risques présents dans le marché financier.

I. LES PRINCIPAUX RISQUES BANCAIRES

I.1. Risque de crédit

Le risque de crédit peut être défini comme le risque de pertes consécutives au défaut d'un emprunteur sur un engagement de remboursement de dettes qu'il a contractées. En général, on distingue trois composantes :

- Le risque de défaut correspond à l'incapacité du débiteur à faire face à ses obligations.
- La deuxième composante du risque de crédit provient de l'incertitude pesant sur le taux de recouvrement une fois le défaut survenu.

La dégradation de la qualité du crédit constitue la troisième source de risque portant sur une dette. Si la perception de la qualité de l'emprunteur se détériore, la prime de risque accordée par les marchés financiers s'accroît en conséquence. De plus, si l'emprunteur bénéficie d'une note de la part d'une agence de notation, celle-ci est susceptible de se dégrader suite à la perception négative des marchés.

I.2. Risque de marché

Le risque de marché est présenté de la manière suivante dans le texte de la Commission Bancaire française: Le risque de marché, défini comme le risque de pertes sur les positions du bilan et du hors bilan à la suite de variations des prix de marché, recouvre :

- Les risques relatifs aux instruments liés aux taux d'intérêt et titres de propriété du portefeuille de négociation.
- Le risque de change et le risque sur les produits de base encourus pour l'ensemble de l'activité de bilan et hors-bilan.

I.3. Risque opérationnel

Le risque opérationnel a été officiellement défini et pris en compte dans le document soumis à consultation par le comité de Bâle (2001) comme le risque de pertes pouvant résulter de procédures internes inadéquates ou non appliquées des personnes, des systèmes ou d'événements externes. Ces événements de risque sont les fraudes internes ou externes, les risques qui touchent aux relations clients, les problèmes liés à la gestion du personnel, les dommages qui pourraient toucher les actifs physiques, l'interruption totale ou partielle des systèmes ou des processus, et la mauvaise exécution de certains processus qu'ils soient internes ou externes à la banque.

II. PRESENTATION DU BALE I

En 1988 le comité de Bâle met en place le premier accord de Bâle (Accord de Bâle I), appelé également ratio de Cook. Ce ratio a comme objectif de renforcer la solidité et la stabilité du système bancaire international ainsi que d'atténuer les inégalités concurrentielles entre les banques. Basé sur une méthodologie simple, il établissait un minimum d'exigence de couverture des risques de crédit par des fonds propres. Le ratio Cook a été complété par un amendement introduisant le suivi des risques de marché. Ce ratio doit être respecté par l'ensemble des établissements de crédit et les entreprises d'investissement.

En effet, les activités bancaires amènent les banques à prendre des risques qui peuvent générer des pertes. Celles-ci varient dans le temps et en fonction des types d'activités exercées. La banque doit donc disposer de fonds propres suffisants pour couvrir ces pertes et poursuivre son activité.

Le capital réglementaire instauré par l'accord de Bâle I représente le niveau minimum de fonds propres que la banque doit détenir pour assurer la protection des déposants et la stabilité du système financier. Son niveau est fixé par le Comité de Bâle.

Les fonds propres sont constitués des fonds propres de base, les fonds propres complémentaires, et les fonds propres surcomplémentaires.

Où la pondération suit une logique de forfait en fonction de la nature du débiteur, de la localisation du risque, des sûretés reçues et de la durée des engagements.

Cependant l'accord Bâle I comporte un nombre important de faiblesses parmi lesquelles on trouve :

- Les différents degrés de pondération du risque de crédit ne sont pas suffisamment calibrés pour déterminer et différencier adéquatement les emprunteurs.

- Les capacités d'arbitrage des exigences de capital réglementaire ne sont pas considérées (non prise en compte de la titrisation (Technique consistant à sortir des créances du bilan.

Dans les cas de titrisation portant sur le compte clients, l'entreprise touche du "cash" et améliore son BFR.) et des engagements à moins de 1 an).

- Il n'y a aucun traitement particulier pour les techniques de réduction des risques (dérivés de crédit, accords de compensation).

- Le portefeuille bancaire n'est pas soumis à une mesure sur le risque de taux (fluctuation des taux d'intérêts qui se répercute sur la valeur des actifs à revenu fixe)

- Pas de prise en compte de la complexité grandissante des marchés financiers

III. PRESENTATION DES NOUVELLES RECOMMANDATIONS BALE II

En 1988, le Comité de Bâle I a proposé la mise en place du ratio Cooke, qui impose aux banques de disposer d'un montant de fonds propres proportionnel à leur encours de crédit. Après avoir intégré les risques de marché au ratio Cooke en 1996, le comité de Bâle présidé par W.McDonough en a décidé la refonte en 1999. La logique qui sous-tend cette réforme est simple : elle suggère le passage d'une méthode purement quantitative et forfaitaire à une méthode ajoutant le qualitatif au quantitatif et partant plus sensible à la qualité intrinsèque des risques. Plus précisément, elle vise à réconcilier le capital économique et le capital réglementaire.

Les consultations soumises à la profession bancaire par le comité de Bâle, en vue de la mise en place d'un nouveau ratio de solvabilité McDonough insistent sur les points suivants :

- Une plus grande différenciation dans le traitement des risques de crédits : l'incitation à adopter un nouveau système de notation interne concernant le risque de crédit permettant aux banques d'estimer par elles-mêmes, aux moyens de leurs informations internes, la charge en capital, c'est à dire le montant des fonds propres nécessaires pour couvrir ce risque de crédit.

Pour ce faire, la banque note ses clients à partir des informations recueillies en internes et publiques, pour les classer ensuite en portefeuilles homogènes. Cette note dérivera du calcul de la perte attendue définie comme étant le produit de la probabilité de défaut (qui sera estimée par la banque), la perte en cas de défaut et de l'exposition au moment du défaut.

- L'importance de la discipline de marché reposant sur la communication régulière d'informations par la banque au marché, ce qui accentue son pouvoir de contrôle et de sanction des banques. La diffusion d'informations significatives par les banques apporte des éléments aux intervenants et facilite l'exercice d'une discipline de marché efficace. Une amélioration de la transparence présentera des avantages pour les banques bien gérées, les investisseurs et les déposants ainsi que pour le système financier d'une manière générale pour éviter le risque systémique.

Trois piliers indissociables sont à la base de ce nouveau dispositif qui renforce les synergies entre contrôle interne et externe des risques (Figure 13).

III.1. Les exigences minimales en fonds propres

L'ancien ratio minimum de fonds propres ne couvrait que deux types de risques dans la définition des actifs pondérés, le risque de crédit et le risque de marché. L'accord Bâle II comporte des changements substantiels du traitement du risque de crédit, le maintien inchangé du dispositif de 1996 sur le risque de marché et l'introduction explicite du risque opérationnel qui conduit à inclure une mesure de ce risque au dénominateur du ratio de fonds propres d'une banque.

Le risque opérationnel est défini comme étant : « le risque de perte résultant de l'inadaptation ou la défaillance de procédures, de personnes ou de système, ou d'événements extérieurs ».

Un aspect novateur du premier pilier est la sensibilité au risque de crédit de la dotation en fonds propres qui tient aussi compte de la solvabilité de l'emprunteur et de la qualité des garanties qu'il apporte. En premier lieu, les banques ont le choix entre trois options pour déterminer le niveau de fonds propres requis ; elles peuvent utiliser d'une part les évaluations externes d'agences de notation reconnues (approche standardisée) et d'autre part leurs modèles de notation interne (approche NI) en adoptant une approche simple ou complexe. En deuxième lieu, les exigences minimales de fonds propres dépendent de la catégorie de contrepartie correspondante et de ses risques spécifiques : grandes entreprises, Etats et administrations publiques, banques, particuliers et petites entreprises, et participations.

III.2. Processus de surveillance prudentielle

L'objectif principal du deuxième pilier est de s'assurer que les banques évaluent au mieux l'adéquation de leurs fonds propres au regard de leur profil de risque.

Premièrement il incite les professionnels bancaires à développer leur techniques de gestion et de contrôle de l'ensemble des risques et ainsi à mettre en place des procédures d'évaluation du capital économique nécessaire pour couvrir l'ensemble de ces risques. Le capital économique représente les fonds propres que la banque doit détenir pour absorber les pertes exceptionnelles. Il est nécessaire de le calculer pour le rapprocher au capital réglementaire.

Dans un second temps il définit les prérogatives du comité de Bâle en matière de contrôle et le cas échéant, ses possibilités d'intervention sur les exigences et les processus de gestion des risques afin de garantir le respect des normes. De plus il incite les autorités de contrôle à adapter les exigences minimales en fonds propres de chaque établissement en fonction de son profil et de la qualité de leur suivi.

III.3. Discipline de marché

La diffusion disciplinée et rigoureuse d'informations au marché constitue le troisième pilier. En communiquant des informations détaillées sur tous les types de risques, une banque permet à tous les autres acteurs du marché de mieux analyser son profil de risque et l'adéquation de ses fonds propres.

Chapitre3 : Définition du Défaut et du processus de recouvrement

1- Définition du Défaut

On dit Défaut lorsque :

« L'établissement estime improbable que le débiteur rembourse en totalité son crédit sans qu'il y ait besoin de prendre des mesures appropriées telles que la réalisation d'une garantie »

OU / ET

« L'arrière du débiteur sur l'un de ses engagements significatifs dépasse 90 jours... »

D'autres signes de défaut :

« L'établissement attribue à une exposition le statut de CES ou créance irrégulière... »

« L'établissement comptabilise une annulation ou constitue une provision après avoir constaté une détérioration significative de la qualité de crédit par rapport à l'exposition »

« L'établissement cède la créance en enregistrant une perte économique significative »

« L'établissement autorise la restructuration en urgence de la créance impliquant vraisemblablement une réduction de l'engagement financier... »

Horizon de défaut :

C'est la maturité du crédit ou la maturité moyenne du portefeuille, dans notre étude on va se contenter de l'horizon d'un an.

2- Processus de recouvrement



Projet fin d'étude
2006-2007

Présentation de La Base de Donnée

Nous allons essayer de présenter durant cette partie les différentes étapes de construction de la base de données, l'échantillon de notre étude ainsi que les caractéristiques de notre portefeuille d'étude, à travers des études descriptives qui montrent la répartition des clients de notre portefeuille selon les différents critères retenus pour l'étude.

I. ELABORATION DE LA BASE DE DONNEES

I.1. Définition du portefeuille d'étude

Concernant le portefeuille du crédit agricole, il est constitué de quatre marchés à savoir : petites et moyennes entreprises agricoles (PMEA), filière agricole (FA), petites et moyennes industries, petites et moyennes entreprises professionnelles (PMI et PME professionnelles) et enfin le marché corpo (constitué des entreprises agricoles dont le chiffre d'affaires est très élevé).

En ce qui concerne notre marché sur lequel on a abordé notre travail, c'est le marché PMI et PME professionnelles et ce, pour plusieurs raisons : les autres marchés sont déjà notés par une société extérieure et la disponibilité partielle des données. De plus, Ce marché représente 20% du portefeuille du CAM.

I.2. les étapes de la construction de la base de données

- Première phase

Dans la première phase de la construction de la base de données, on a réservé presque une semaine pour réfléchir aux variables qui peuvent être intéressantes dans notre étude, que ce soit des variables quantitatives de types ratios financiers qui reflètent l'état financier de l'entreprise ou des variables qualitatives qui reflètent la manière d'organisation ou de management de l'entreprise et aussi le secteur d'activité, et ceci en discussion avec nos encadrants. Cette phase nous a permis de faire sortir des critères qui s'avèrent nécessaires à l'évaluation du portefeuille, comme le montre bien le tableau ci-dessous :

Critères et sous critères pour l'évaluation de la qualité du portefeuille bancaire du CAM
--

Chaque information pour l'analyse du risque à posteriori a été décrite à partir du canevas suivant :

Quoi	Préciser l'information
Où	Où est l'information ?
Comment	Comment l'obtenir (élément permettant de la calculer)
Pourquoi	à quoi sert cette information ? pour quels enjeux ?

A. Secteur d'activité

A.1. Tendances du secteur

Quoi	Tendance du secteur
Où	Rapport de gestion, études sectorielles
comment	Choix parmi la liste suivante : <ul style="list-style-type: none"> ▪ En croissance ▪ Stable ▪ En déclin
pourquoi	Indique si le marché est porteur ou pas

A.2. Nature du secteur

quoi	Nature du secteur
Où	Rapport de gestion, études sectorielles
Comment	Secteur monopolistique, oligopolistique ou concurrentiel : <ul style="list-style-type: none"> ▪ Nombre d'entreprise réalisant plus de 10% du CA du secteur ▪ Poids des principales entreprises (somme des CA des 3 premiers entreprises / CA du secteur)
pourquoi	

B. Structure financière

S'assurer de la certification des comptes et relever les éventuelles réserves

B.1. Solvabilité

Quoi	Solvabilité
Où	Comptes de l'entreprise
Comment	Fonds propres / Total bilan
pourquoi	

B.2. Rentabilité

Quoi	Rentabilité (rentabilité d'exploitation et rentabilité financière)
Où	Comptes de l'entreprise
Comment	Rentabilité d'exploitation : EBE / CA HT Rentabilité financière : Résultat net / capitaux propres
pourquoi	

B.3. Service de la dette

Quoi	Service de la dette
Où	Comptes de l'entreprise
Comment	<ul style="list-style-type: none"> ▪ EBE / frais financiers ▪ Capacité d'autofinancement / Dette
pourquoi	

B.4. Valeur relative du fonds de roulement net global

Quoi	Valeur relative du fonds de roulement net global
Où	Comptes de l'entreprise
Comment	$(FRNG / CA HT) * 360$
Pourquoi	Indique d'importance du fonds de roulement

C. Atouts de la relation

C.1. Ancienneté dans le secteur

Quoi	Ancienneté dans le secteur
Où	
Comment	Nombre d'années de l'entreprise dans le secteur
Pourquoi	

C.2. politique d'investissement

Quoi	Qualité de l'outil de production
Où	
Comment	Immobilisations, degré d'amortissement, politique de renouvellement
Pourquoi	Maintien de la compétitivité, développement

D. Relations bancaires

D.1. Nombres de banques en relation avec le client

Quoi	Nombre de banques en relation avec le client
Où	
Comment	Nombre de banques, liste de banques, nombre de comptes bancaires
pourquoi	

D.2. part financement CAM

Quoi	Part financement CAM
Où	Comptes, SI banque
Comment	Dettes CAM / dettes bancaires
pourquoi	

D.3. Poids de l'impayé

Quoi	Poids de l'impayé
Où	SI banque
Comment	<ul style="list-style-type: none"> ▪ Montants impayés / Encours total ▪ Fréquence des impayés
pourquoi	

- Deuxième phase

Après la détermination des critères qui s'avèrent nécessaires pour évaluer la qualité du portefeuille, nous avons passé à l'étape de la consultation des données disponibles concernant les différentes variables où nous avons exclu pas mal de critères à cause de l'indisponibilité des informations concernant notre portefeuille.

- Troisième phase

Cette phase a été caractérisée par une réunion où nous avons rencontré les chefs de service des départements audit et qualité de risque et département des études stratégiques. Pendant cette phase, nous avons défini toutes les variables proposées ainsi que la variable correspondante à la qualité du client « bon » ou « défaut », enfin notre base de données est de 35 entreprises qui construisent le portefeuille réparti sur 4 ans d'historique c'est à dire 141 observations et un horizon de travail d'un an.

- Quatrième phase

Nous avons calculé certains ratios financiers à partir des états de synthèse des entreprises qui construisent notre portefeuille d'étude à savoir les ratios de solvabilité, ratios de rentabilité d'exploitation, ratios de rentabilité financière, fonds de roulement net global,...

I.3. Les variables retenues pour l'étude

Parmi toutes les variables déjà exposées précédemment, nous allons nous contenter d'utiliser 12 descripteurs, et cela est dû d'une part du fait que les ratios dont nous disposons, ne sont pas tous calculés (bilan illisible, valeurs manquantes) et d'autre part, pour éviter le problème de la redondance, surtout pour certains ratios financiers, comme il est expliqué dans les trois phases précédentes.

Les variables finales sont :

- solv (R1) : Ratio de solvabilité de l'entreprise
- rentexploit (R2) : Ratio de rentabilité d'exploitation
- rentfinanc (R3) : Ratio de rentabilité financière
- servidette : service de la dette
- frngca : fonds de roulement net global sur le chiffre d'affaire (par jours)
- chargpersonel : charges de personnels
- anciénte : l'ancienneté de l'entreprise dans le secteur
- montecheance : le montant de l'échéance
- qualsecteur : la qualité du secteur
- interet : la structure du taux d'intérêt sollicité
- ratio12 : montant demandé / dette
- partfinanc : la part financée par le CAM dans le portefeuille de l'entreprise.

II. ETUDE DESCRIPTIVE

II.1. Statistique univariées

Nous avons jugé nécessaire de disposer des statistiques univariées pour chaque groupe étudié concernant les clients bons et défauts sur les variables quantitatives considérées, et ceci pour avoir une idée globale et préalable sur le comportement de ces deux groupes vis à vis des descripteurs en question, comme le présente le tableau1 ci-dessous, qui donne les moyennes et les variances pour les deux catégories des clients selon les variables quantitatives retenues pour l'étude, et qui sont calculées par le logiciel SPSS.

Tableau1 : Caractéristiques des clients en défaut

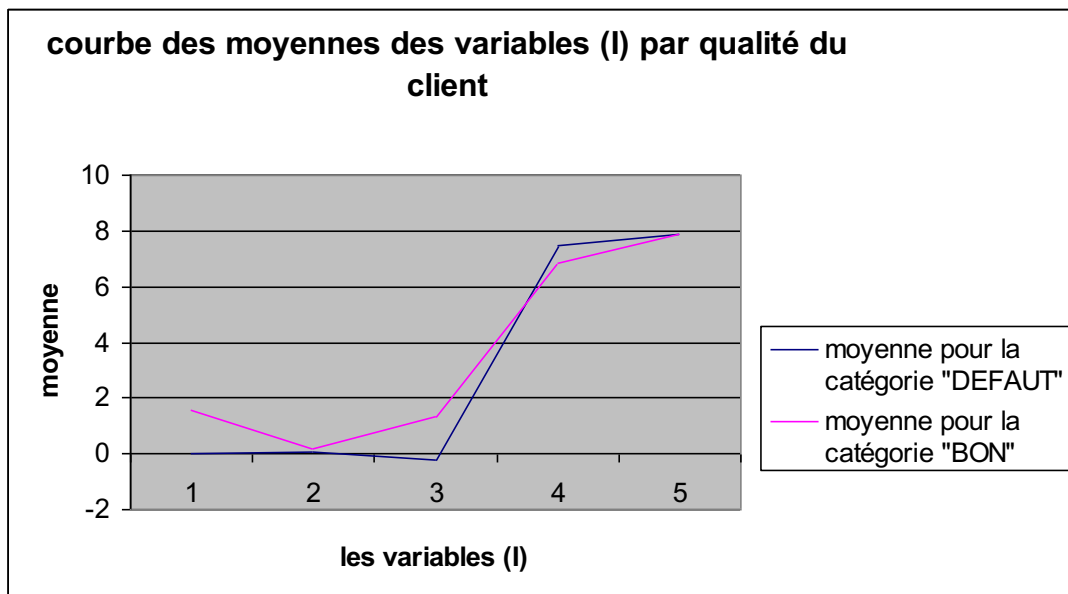
<i>DEFAUT</i>	N	Moyenne	Ecart type	Variance
solva (R1)	73	0,02060143	0,15558473	0,02420661
rentexp (R2)	73	0,04477422	0,06059344	0,00367157
rentfin (R3)	73	-0,21538421	1,59978797	2,55932156
servdett (R6)	73	38,7509698	133,482721	17817,6367
FRNG/CA(R4)	73	-5,79027185	275,999331	76175,6305
chargper(R5)	73	2075076,82	3758265,05	1,4125E+13
annciété	73	7,46575342	4,2690169	18,2245053
INTERET	73	7,86209631	1,86021246	3,4603904
montechean	73	27051,8462	4837,56891	23402073
RATIO12	73	143,031359	130,178709	16946,4962

Tableau2 : Caractéristiques des clients bons

BON	N	Moyenne	Ecart type	Variance
solva (R1)	68	1,53124987	10,3495287	107,112744
rentexp (R2)	68	0,1554449	0,4803122	0,23069981
rentfin (R3)	68	1,32705506	7,10601882	50,4955035
servdett (R6)	68	781,735988	4164,43522	17342520,7
FRNG/CA(R4)	68	35,6185667	353,587347	125024,012
chargper(R5)	68	5434787,94	13733794,5	1,8862E+14
anncienté	68	6,86764706	4,46179273	19,9075944
INTERET	68	7,88113536	1,63740646	2,6810999
montechean	68	26729,0229	4591,68973	21083614,5
RATIO12	68	116,037626	119,410293	14258,8182

D'après ces deux tableaux présentés ci-dessus, on remarque que la moyenne de ratio de solvabilité pour les deux catégories de clients est bien assez différente ce qui montre que ce critère discrimine bien les deux groupes, ceci est aussi remarqué pour le ratio de rentabilité financière et service de la dette, alors que les autres critères ont des moyennes proches pour les deux groupes de clients, tandis que la variance correspondante aux différents pour les deux types de clients est inégales en générale, comme le montre bien le graphique ci-dessous :

Graphique16 : Courbes de moyenne par groupe de clients



On déduit d'après le graphique ci-dessus que les deux variables 1 et 3 qui correspondent respectivement au ratio de solvabilité et de rentabilité financière, ont des valeurs différentes pour les deux catégories de clients, alors que les autres variables ont des valeurs proches, et cela est aussi valable pour le variable 2 qui correspond au service de la dette (voir graphique17 annexe II), ce qui va être vérifié par la suite à travers les tests d'égalités de la variance et de la moyenne, comme il est présenté dans le tableau ci-dessous calculé par SPSS :

Tableau : tests d'égalité de la moyenne et de la variance pour les deux types de clients

	F ²	Sig.	t	ddl	Sig.(bila)	Diffé moye	Diff écart	Borinf	Borsup
solva (R1)	4,2137942	0,0419724	-1,247287	139	0,21439	-1,510648	1,21115	-3,905	0,884
rentfin (R3)	3,3660042	0,0686931	-1,806515	139	0,073001	-1,542439	0,85382	-3,231	0,1457
servdett (R6)	8,605748	0,003921	-1,523915	139	0,129802	-742,985	487,55	-1707	220,99

2 : Test de Levene sur l'égalité des variances

On constate selon le tableau ci-dessus que l'hypothèse d'égalité de la variance est rejetée pour les deux variables explicatives, le ratio de solvabilité et service de la dette, puisque la significativité du test est inférieure à 5%, alors qu'elle est acceptée pour la variable rentabilité financière. Tandis que l'hypothèse d'égalité de la moyenne pour les deux groupes de clients concernant les trois variables, le ratio de rentabilité, rentabilité financière et service de la dette, est acceptée.

II.2. Les modalités des critères retenus

Afin d'utiliser seulement des variables qualitatives, nous avons discrétisé les variables quantitatives en classes de catégories dont le nombre dépend du type des variables quantitatives et aussi du nombre d'observations ou d'individus dans chaque classe.

- Ancienneté

<i>Ancient</i>	Inférieur à 3ans	Entre 3ans et 8ans	Supérieur à 8ans
<i>code</i>	1	2	3
<i>Pourcentage</i>	30%	27%	43%

- Intérêt

<i>intérêt</i>	Inférieur à 7%	Entre 7% et 8,5%	Supérieur à 8,5%
<i>Code</i>	1	2	3
<i>pourcentage</i>	39%	22%	39%

- Solvabilité

<i>Solva (R1)</i>	Inférieur à 0	Entre 0 et 0,20	Supérieur à 0,20
<i>Code</i>	1	2	3
<i>pourcentage</i>	8,5%	62%	29,5%

- Rentabilité d'exploitation

<i>Rent d'exploit (R2)</i>	Inférieur à 0,035	entre 0,035 et 0,12	Supérieur à 0,12
<i>Code</i>	1	2	3
<i>pourcentage</i>	46%	36%	18%

- Rentabilité financière

<i>Rentfin (R3)</i>	Inférieur à 0,01	entre 0,01 et 0,33	Supérieur à 0,33
<i>Code</i>	1	2	3
<i>pourcentage</i>	44%	37%	19%

- Montant échéance (en DH)

<i>montéchéance</i>	Inférieur à 20000	entre 20000 et 28000	Supérieur à 28000
<i>Code</i>	1	2	3
<i>pourcentage</i>	11%	35%	54%

- Ratios12 (montant accordé / dette)

<i>Ratios12</i>	Inférieur à 100	entre 100 et 170	Supérieur à 170
<i>Code</i>	1	2	3
<i>pourcentage</i>	48%	19%	32%

- Charges personnelles (dh)

<i>chargepersonl</i>	Inférieur à 10 ⁶	Entre 10 ⁶ et 3*10 ⁶	Supérieur à 3*10 ⁶
<i>Code</i>	1	2	3
<i>pourcentage</i>	53%	23%	24%

- Fonds de roulement net global sur chiffre d'affaire (en jours)

<i>frngca</i>	Inférieur à 10	Entre 10 et 50	Supérieur à 50
<i>Code</i>	1	2	3
<i>pourcentage</i>	42%	23%	35%

- Service de la dette

<i>servicedette</i>	Inférieur à 0	Entre 0 et 50	Supérieur à 50
<i>Code</i>	1	2	3
<i>pourcentage</i>	39%	41%	20%

- Qualité du secteur

<i>qualitésect</i>	En Déclin	Stable	En Croissance
<i>Code</i>	1	2	3
<i>pourcentage</i>	15%	45%	40%

II.3. Répartition des clients par modalité

Afin d'avoir une idée sur le comportement ou la qualité des clients vis-à-vis des différentes variables retenues pour l'étude, nous avons construit un tableau qui montre la répartition des deux catégories de clients pour les différentes modalités correspondantes aux variables d'étude, comme l'explique bien le tableau ci-dessous :

Tableau : Répartition des clients par modalité

	<i>BON</i>	<i>DEFAULT</i>
modalités	Pourcentage%	Pourcentage%
sol_1	8	92
sol_2	30	70
sol_3	100	0
rentexp_1	47	53
rentexp_2	37	63
rentexp_3	75	25
rentfin_1	43	57
rentfin_2	51	49
rentfin_3	56	44
serdett_1	55	45
serdett_2	39	61
serdett_3	51	46
frng_1	45	55
frng_2	42	58
frng_3	56	44
chargper_1	41	59
chargper_2	50	50
chargper_3	62	38
ancint_1	51	49
ancint_2	50	50
ancint_3	45	55
qualsec_1	49	51
qualsec_2	55	45
qualsec_3	42	58
interet_1	40	60
interet_2	64	36
interet_3	47	53
partfin_1	45	55
partfin_2	53	47
partfin_3	46	54
montech_1	53	47
montech_2	47	53
montech_3	48	52
ratios12_1	54	46
ratios12_2	39	61
ratios12_3	46	54

D'après le tableau ci-dessus, on remarque que les clients bons sont caractérisés ou majoritaires pour les modalités suivantes : solv_3, rentexpl_3, c'est-à-dire la population des clients qui ont un ratio de solvabilité supérieur à 0,2, un ratio de rentabilité d'exploitation supérieur à 0,12 et un taux d'intérêt sollicité entre 7 et 8,5%. Alors que les clients en défaut ont une part grande pour la modalité sol_1, caractérisée par un ratio de solvabilité inférieur à zéro, comme il y a des modalités où les pourcentages correspondants aux deux groupes de clients sont presque égaux, par exemple les modalités rentexp_1, rentfin_2, ancienneté_2, ce qui veut dire qu'elles ne peuvent pas être des modalités discriminantes.

II.4. Test des hypothèses de normalité et d'indépendance des critères

I.4.1. La normalité des variables quantitatives

Nous allons tester l'hypothèse de la normalité des descripteurs quantitatifs au seuil de 5%, en utilisant la statistique de test de Shapiro-Wilk calculé par le logiciel XL-stat, comme il est montré dans le tableau ci-dessous :

H_0 : l'échantillon suit une loi normale vs H_1 : l'échantillon ne suit pas une loi normale

Tableau5 : significativité de la normalité des variables

variables	W(valeur observé)	p-value	signification
solvabilité	0,072	< 0,0001	0,05
rentaexploita	0,258	< 0,0001	0,05
rentfinancier	0,206	< 0,0001	0,05
service dette	0,118	< 0,0001	0,05
FRNG/CA	0,654	< 0,0001	0,05
chargesper	0,367	< 0,0001	0,05

Conclusion :

Au seuil de signification Alpha=0,05 on rejette l'hypothèse nulle selon laquelle l'échantillon suit une loi normale pour toutes les variables explicatives, puisque toutes les p-values sont inférieures au seuil fixé, autrement dit, la non normalité des ratios financiers présentés ci-dessous est significative au seuil de 5%.

I.4.2. La dépendance des variables descriptrices avec la variable expliquée

Nous allons utiliser le test de khi deux, fournit par le logiciel statistique XL-stat, afin de tester l'hypothèse de l'indépendance de la variable expliquée (variable dichotomique) avec les différentes variables explicatives, comme il est présenté dans le tableau ci-dessous :

H_0 : il y a indépendance entre la variable expliquée Y et le descripteur X

Tableau6 : Significativité de l'indépendance des variables

Variables	Khi-deux	ddl	signification
Solvabilité (R1)	3,91	138	1,00
Rentabilité expl (R2)	742,90	109	,00
Rentabilité fin (R3)	23,93	135	1,00
Service dette (R6)	784,34	108	,00
FRNG/CA (R4)	,00	140	1,00
Charge personel (R5)	473,87	116	,00
Anncienté	37,82	13	,00
Qualité secteur	,80	2	,66
Intérêt	,00	140	1,00
Partfinan CAM	4,38	2	,112
Montant échéance	9,723	131	1,00
Ratios 12	2,87	137	1,00

Les valeurs de khi-deux pour les différentes variables représentées dans le tableau ci-dessus, indiquent que l'hypothèse d'indépendance est rejetée au seuil de signification 5%, sauf pour les deux variables FRNG/CA et intérêt dont les valeurs de khi-deux sont inférieures

au seuil fixé. Donc on peut déduire qu'il y a une dépendance en général entre la variable expliquée et les variables explicatives choisies et que les deux variables FRNG/CA et taux d'intérêt ne sont pas des facteurs de discrimination.

Conclusion

D'après l'étude de la statistique univariée, il s'avère que le ratio de solvabilité peut être un facteur de discrimination pour les deux catégories du clients, et selon le test de khi deux, nous avons constaté que le ratio FRNG/CA ainsi que le taux d'intérêt sont indépendants de la variable expliquée, c a dire que ces deux variables ne sont pas qualifiées d'être des facteurs de discrimination. Ce que nous allons vérifier par la suite à travers les deux méthodes de discrimination, Disqual et la Régression Logistique.

Projet fin d'étude
2006-2007

APPROCHE THEORIQUE :

- ***Méthode Disqual***
- ***Régression Logistique***
- ***Validation du modèle***

Chapitre 1: Fondement théorique de la méthode DISQUAL

La méthode Disqual (discrimination sur variables qualitatives) est un autre exemple de méthode bien adaptée aux variables qualitatives. Mais cette fois les données doivent toutes être des variables à modalités. Elle a été mise au point par J-M. Bourroche et G. Saporta. Dans les années 80, le Crédit Commercial de France a développé cette méthode pour la détection précoce des difficultés d'entreprise en utilisant des ratios économiques et financiers transformés en variables qualitatives par codage.

Cette méthode présente le grand avantage de prendre en compte les non linéarités de certaines variables par rapport au phénomène étudié. Elle permet aussi de renforcer le pouvoir discriminant des descripteurs en les croisant. La phase de préparation des variables est donc délicate : au choix des variables elles-mêmes, qualitatives ou quantitatives, s'ajoute la difficulté de définir efficacement les répartitions en classes des variables quantitatives pour les coder.

Ensuite, viendra le déroulement de la méthode elle-même dont on verra qu'elle présente l'avantage de ne pas nécessiter d'hypothèse sur les lois suivies par les variables. La perte d'information due à la répartition en classes de variables quantitatives n'est pas qu'apparente, puisqu'elle permet ensuite la prise en compte de phénomènes non linéaires.

Déroulement des procédures statistiques

Le tableau des données codées est transformé en tableau disjonctif complet grâce aux variables indicatrices des modalités.

S'il y a p variables, chacune a h_j modalités. Le nouveau tableau de données correspond aux différentes modalités a pour terme général m_{ih} ou i est l'indice de l'individu et h l'indice de la modalité.

L'analyse des correspondances multiples (ACM) sur ce tableau disjonctif complet définit des axes factoriels. Les colinéarités sont éliminées par la diagonalisation qui réduit la dimension de l'espace des variables. Le nuage de points est représenté par ses coordonnées sur les axes factoriels.

Sur ce nouveau tableau des coordonnées sur les axes les plus discriminants, est pratiquée une analyse discriminante linéaire de Fisher (AFD), le tableau des nouvelles données ainsi régularisées satisfaisant mieux les hypothèses optimales de l'analyse de Fisher.

I. ANALYSE DES CORRESPONDANCES MULTIPLE

L'analyse des correspondances multiples permet de décrire de vastes tableaux binaires dont les fichiers d'enquêtes constituent un exemple édifiant : les lignes de ces tableaux sont en général des individus, observations, ou objets (dizaines de milliers) ; les colonnes sont des modalités de variables qualitatives, le plus souvent des modalités de réponses à des questions. Il s'agit d'une extension de l'analyse des correspondances simple (AFC) avec cependant des procédures de calcul et des règles d'interprétation spécifiques.

I.1. Tableau disjonctif complet

On désigne par I l'ensemble de n sujets ayant répondu au questionnaire et par p le nombre total des modalités des s questions.

Chaque question q possède p_q modalités : $q = 1, 2, \dots, s$

Avec : $p = \sum p_q$ $q = 1, 2, \dots, s$

On note \mathbf{Z} le tableau à n lignes et p colonnes décrivant les s réponses des n individus par un codage binaire, c'est une juxtaposition de s sous-tableaux où chacun correspond à une question :

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_s]$$

Le tableau \mathbf{Z} est appelé tableau disjonctif complet dont le terme général s'écrit :

$Z_{ij} = 1$ si l'individu i choisit la modalité j de la question q

$Z_{ij} = 0$ autrement

Les marges en ligne du tableau disjonctif complet sont constantes et égales au nombre s de questions :

$$Z_{i.} = \sum Z_{ij} = s$$

Les marges en colonne : $Z_{.j} = \sum Z_{ij}$ correspondent au nombre de sujets ayant choisi la modalité j de la question q .

On vérifie que, pour chaque sous-tableau \mathbf{Z}_q , l'effectif total est : $\mathbf{z}_q = \sum \mathbf{z}_{.j} = \mathbf{n}$

La somme des marges donne l'effectif total z du tableau \mathbf{Z} soit : $\mathbf{Z} = \sum \sum \mathbf{z}_{ij} = \mathbf{ns}$

I.2. Principes de l'analyse des correspondances multiples

Les principes de cette analyse sont les mêmes que l'AFC c'est-à-dire :

- Mêmes transformations de tableaux de données ;
- Même critère d'ajustement ;
- Même distance, celle du χ^2 Khi deux.

L'analyse des correspondances multiples présente par ailleurs des propriétés particulières dues à la nature du tableau disjonctif complet.

I.2.1. Critère d'ajustement et distance du χ^2

Les individus sont tous affectés d'une masse identique m_i avec : $m_i = 1/n$ $\forall i = 1, \dots, n$

Chacune des modalités j est pondérée par sa fréquence $m_j = Z_{.j}/ns$

La distance, dans R^n , entre deux modalités j et j' est la suivante :

$$d^2(j, j') = \sum_{i=1}^n n \left(\frac{Z_{ij}}{Z_{.j}} - \frac{Z_{ij'}}{Z_{.j'}} \right)^2$$

Deux modalités choisies par les mêmes individus coïncident. Par ailleurs, les modalités de faible effectif sont éloignées des autres modalités.

La distance, dans R^p , entre deux individus i et i' est la suivante :

$$d^2(i, i') = \frac{1}{s} \sum_{j=1}^p \frac{n}{Z_{.j}} (z_{ij} - z_{i'j})^2$$

Deux individus sont proches s'ils ont choisi les mêmes modalités. Ils sont éloignés s'ils n'ont pas répondu de la même manière.

I.2.2. Axes factoriels et facteurs

La matrice X à diagonaliser est de terme général : $x_{jj'} = \frac{1}{s \cdot z_{.j'}} \sum_{i=1}^n z_{ij} \cdot z_{ij'}$

Dans R^p , le $\alpha^{\text{ème}}$ axe factoriel est engendré par le vecteur propre u_α associé à la $\alpha^{\text{ème}}$ plus grande valeur propre de la matrice X.

Si on note C_α le $\alpha^{\text{ème}}$ facteur dans R^p et D_α le $\alpha^{\text{ème}}$ facteur dans R^n , on aura :
 $XC_\alpha = \lambda_\alpha C_\alpha$ et $X'D_\alpha = \lambda_\alpha D_\alpha$.

I.2.3. Facteurs et relations quasi-barycentriques

La coordonnée factorielle de l'individu i sur l'axe α est donnée par :

$$C_{\alpha i} = \frac{1}{\sqrt{\lambda_\alpha}} \cdot \sum_{j=1}^p \frac{z_{ij}}{z_{.i}} \cdot D_{\alpha j}$$

Par construction du tableau disjonctif complet, il en découle :

$$C_{\alpha i} = \frac{1}{s \cdot \sqrt{\lambda_\alpha}} \cdot \sum_{j \in P_i} D_{\alpha j}$$

P_i désigne l'ensemble des modalités caractérisant l'individu i.

La coordonnée factorielle de la modalité j sur l'axe α est donnée par :

$$D_{\alpha j} = \frac{1}{\sqrt{\lambda_\alpha}} \cdot \sum_{i=1}^n \frac{z_{ij}}{z_{.j}} \cdot C_{\alpha i}$$

C'est-à-dire :

$$D_{\alpha j} = \frac{1}{z_{.j} \sqrt{\lambda_\alpha}} \cdot \sum_{i \in Q_j} C_{\alpha i}$$

Q_j désigne l'ensemble des individus caractérisés par la modalité j.

Interprétation

Au coefficient $1/\sqrt{\lambda_\alpha}$ près, l'individu i se trouve au point moyen du nuage des modalités qu'il a choisies.

- La modalité j se trouve au point moyen du nuage des individus qui l'ont choisie comme réponse.

La représentation simultanée des individus et des modalités est importante pour l'interprétation des résultats. Mais non utilisée dans la pratique car, dans le cas de plusieurs dizaines de milliers d'individus -comme dans le cadre de notre application- on encombre le graphique de la façon qu'aucune interprétation n'est possible d'autant plus que les individus sont anonymes et ne sont intéressants qu'à travers leurs caractéristiques. On procède alors par la projection du barycentre d'un groupe d'individus homogène sur un plan factoriel afin d'apprécier ses caractéristiques.

I.2.4. Inertie du nuage des modalités et conséquences pratiques

La distance entre la modalité j et le centre de gravité du nuage G vaut :

$$d^2(j, G) = n \sum_{i=1}^n \left(\frac{z_{ij}}{z_{.j}} - \frac{1}{n} \right)^2 = \frac{n}{z_{.j}} - 1$$

Résultat : la distance d'une modalité au centre de gravité du nuage est d'autant plus grande que l'effectif des individus ayant choisi cette modalité est faible.

- Inertie d'une modalité

L'inertie I (j) d'une modalité j vaut :

$$I(j) = 1/s (1 - z_{.j}/n)$$

En voie de conséquence : la part d'inertie due à une modalité de réponse est d'autant plus grande que l'effectif dans cette modalité est plus faible.

Ainsi, à l'occasion de la discrétisation d'une variable continue on évite les classes à faible effectif ayant l'inconvénient de perturber les directions des premiers axes factoriels, d'où la construction de classes équilibrées en termes d'effectifs s'avère nécessaire.

- Inertie d'une question

L'inertie de la question q, notée I (q) vaut :

$$I(q) = \sum_{j=1}^{p_q} I(j) = 1/s (p_q - 1)$$

Ainsi la part d'une d'inertie due à une question est fonction croissante du nombre de modalités de réponse. D'où l'intérêt d'avoir un nombre équilibré de modalité pour toutes les questions afin de leur faire jouer le même rôle.

- Inertie totale

L'inertie totale de toute les modalités est égale à : $p/s - 1$

L'inertie totale dépend uniquement du nombre de variables et de modalités et non des liaisons entre les variables, c'est une quantité qui n'a pas de signification statistique.

I.2.5. Règles d'interprétation

Les règles d'interprétation des résultats sont sensiblement les mêmes qu'en AFC, on calcule la contribution et la qualité de représentation de chaque modalité et de chaque individu si ceux-ci ne sont pas anonymes pour l'analyse.

Par ailleurs, la notion de variable doit être prise en compte au moment de l'interprétation, à travers ses modalités.

La contribution d'une variable nominale au facteur α est la somme des contributions de ses modalités sur ce facteur :

$$Cr_{\alpha}(q) = \sum_{j \in q} Cr_{\alpha}(j)$$

On obtient ainsi un indicateur de liaison entre la variable et le facteur.

II. ANALYSE FACTORIELLE DISCRIMINANTE

II.1. Introduction

On désigne sous le nom d'Analyse Factorielle Discriminante (AFD) une famille de techniques destinées à classer (c'est-à-dire affecter à des classes préexistantes) des individus caractérisés par un certain nombre de variables numériques. Elle permet d'étudier le lien

existant entre des variables quantitatives et un caractère qualitatif. Ce dernier permet de répartir la population étudiée en plusieurs classes.

II.2. Objectifs de l'AFD

L'analyse factorielle discriminante a deux principaux objectifs :

II.2.1. Objectif descriptif

Consiste en la recherche des combinaisons linéaires des variables qui permettent de séparer au mieux les k classes d'individus définies préalablement dans la population, et donner une représentation graphique adéquate qui rende compte de cette séparation.

II.2.2. Objectif prévisionnel

C'est-à-dire prévoir la classe d'affectation d'un nouvel individu décrit par les mêmes variables quantitatives.

II.3. Principe de l'AFD

Comme nous venons de signaler, l'AFD consiste à rechercher les combinaisons de p variables explicatives, généralement continues, qui permettent de séparer au mieux les k classes existantes dans la population.

La première combinaison sera celle dont la variance entre les classes est maximale, afin d'extraire les différences entre les classes, et dont la variance à l'intérieur des classes est minimale pour que l'étendue dans les classes soit délimitée. Puis, parmi les combinaisons linéaires non corrélées à la première, on cherche celle qui discrimine le mieux les classes etc.

II.3.1. Fonction discriminante

On définit une fonction discriminante comme une combinaison linéaire des variables centrées, elle est donnée par :

$$U = X^c u$$

Avec :

X^c : est le tableau des données centrées ;

u : un vecteur de R^p .

La variance de la fonction U est donnée par :

$$V(u) = u' \Gamma u = u'(D+E)u = u'Du + u'Eu$$

Avec :

Γ : Matrice des variances covariances de X ;

D : Variance dans les groupes de la population ;

E : Variance entre groupes de la population.

III.3.2. Calcul des fonctions discriminantes

Une bonne fonction discriminante devrait :

✓ Etre presque constante à l'intérieur d'une classe : $u'Du$ très faible ;

✓ Avoir une grande variabilité entre les classes : $u'Eu$ grande.

Or on a :

$$1 = \frac{u' Du}{u' \Gamma u} + \frac{u' Eu}{u' \Gamma u}$$

Il faut donc maximiser la part $\frac{u' Eu}{u' \Gamma u}$, ce qui équivaut à maximiser $u' Eu$ sous contrainte $u' \Gamma u = 1$.

On peut montrer que le vecteur \mathbf{u} qui permet de donner la meilleure fonction discriminante n'est autre que le vecteur propre de la matrice $\Gamma^{-1}E$ correspondant à la plus grande valeur propre. Cette dernière est dite le **pouvoir discriminant** de la fonction.

Les fonctions discriminantes d'ordre 2,3,... correspondent aux valeurs propre $\lambda_2, \lambda_3, \dots$

Chapitre 2 : Fondements théorique de la Régression Logistique

Un autre modèle fournit une fonction linéaire des descripteurs comme outil de discrimination, c'est la régression logistique. Il repose sur des hypothèses qui généralisent celles de l'analyse discriminante de Fisher dans le cadre multinormal. Cependant il modélise la probabilité à posteriori et non les lois conditionnelles sur les groupes. De plus il permet d'utiliser pour descripteurs des variables binaires et/ou des variables continues. Il présente donc un grand intérêt. Pourtant la comparaison entre l'analyse discriminante de Fisher et la régression logistique est assez délicate et le choix entre ces deux méthodes n'est pas toujours évident.

La préparation des données constitue une phase importante pour l'élaboration d'un tel modèle. En effet, les variables explicatives doivent être arrangées en classes pour bien prélever l'information et afin de bien discriminer entre les groupes. La régression logistique binaire de la variable se fera ensuite sur les modalités des variables.

Afin de construire le meilleur modèle, compte tenu des données, nous utiliserons une méthode de sélection automatique des variables dite méthode descendante : rapport de vraisemblance. Il s'agit de partir d'un modèle de régression contenant toutes les variables, puis à chaque pas effectuer des tests permettant d'éliminer les variables non significatives

Dans le cadre du présent chapitre, on aura l'occasion de présenter dans un premier temps un aperçu sur le fondement théorique et l'hypothèse fondamentale de la discrimination logistique et sur les tests de signification et d'adéquation du modèle pour se lancer après sur l'application du modèle à la notation des clients.

I. PRESENTATION DE LA DISCRIMINATION LOGISTIQUE

I.1. Hypothèse fondamentale de la discrimination logistique

L'hypothèse fondamentale de la discrimination logistique est que le logarithme du rapport des vraisemblances est une fonction linéaire de x :

$$(E) \quad \ln L_1(x)/L_2(x) = \alpha'x + \beta$$

La famille des lois qui vérifient (E) contient :

- les lois multinormales ou $W_1=W_2$
(L'égalité de la matrice variance covariance dans les deux groupes)
- Les lois exponentielles
- Les lois discrètes
- Les lois Gama, Bêta, et les lois de Poison

I.2. Présentation de la régression logistique

Dans la régression logistique binaire (dichotomique), on considère une variable cible binaire $Y=0$ ou 1 , et J variables explicatives X_j continues, binaires ou qualitatives.

L'objectif de la régression logistique est celui de toute régression : modéliser l'espérance conditionnelle $E(Y_i/X=x)$ on veut connaître la valeur moyenne de Y_i pour toute valeur de X_i . Pour une valeur Y_i valant 0 ou 1. Cette valeur moyenne est la probabilité que $Y_i = 1$ (loi de Bernoulli). On a donc :

$$E(Y_i/X=x) = 1 \cdot P(Y_i=1/X=x) + 0 \cdot P(Y_i=0/X=x) = P(Y_i=1/X=x) = p_i$$

En régression linéaire, on cherche à faire passer une droite au milieu du nuage des points (X, Y) , de sorte que l'ensemble des valeurs moyennes de Y pour toutes les valeurs de X est approché par une droite, d'équation :

$$E(Y_i/X=x) = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j$$

Cette approximation ne convient évidemment plus lorsque $Y_i = 0$ ou 1, puisque le terme

$\beta_0 + \beta_1 X_1 + \dots + \beta_j X_j$ est non borné alors que $P(Y_i=1/X=x)$ est dans l'intervalle $[0,1]$.

Le problème posé est qu'une mesure de probabilité est bornée à droite et à gauche, Il convient donc de trouver un moyen de supprimer ces bornes.

En divisant p_i par $(1-p_i)$, la borne à gauche est annulée car lorsque p_i tend vers 1, $p_i/(1-p_i)$ tend vers $+\infty$. Si p_i tend vers 0 alors $p_i/(1-p_i)$ tend aussi vers zéro. On applique alors une seconde transformation, de type logarithmique, de telle manière à ce que lorsque p_i tend vers 0, la transformation tend vers moins l'infini. Cette double transformation est appelée le Logit de p_i .

$$\text{Logit} = \ln(p_i/1-p_i)$$

Le modèle que l'on va donc chercher à estimer, prend donc la forme suivante :

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta X_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_j X_{ji} = Z$$

où $\beta = (\beta_0, \dots, \beta_j)$ sont les paramètres inconnus du modèle qu'il estime.

En manipulant de manière adéquate le modèle précédent on obtient un modèle qui exprime p_i en fonction de z :

$$p_i = P(Y_i=1/X_i=x) = \frac{e^{\beta X_i}}{1 + e^{\beta X_i}}$$

La première forme est la linéarisation de cette seconde expression.

Variables explicatives qualitatives

Dans le cas d'une variable qualitative à modalités non ordonnées, on va introduire une variable indicatrice par modalité.

Exemple : CSP = 1 ou 2 ou 4 ou 5

On définit $CSP_i = 1$ si $CSP=i$

$= 0$ sinon

Cependant, on ne pourra pas garder toutes les modalités ou les cinq variables indicatrices, car elles ne sont pas indépendantes entre elles. En effet, chaque personne a une et une seule CSP, et donc :

$$CSP_1 + CSP_2 + CSP_3 + CSP_4 + CSP_5 = 1$$

D'habitude, la variable représentant la situation la plus courante sert de **situation de référence**, et on la supprime. Cela revient à dire que son coefficient est nul. Dans une régression logistique sur variables binaires, l'ensemble des situations de référence est représenté par la constante.

Variable quantitative

Si la variable R est un descripteur lié linéairement au phénomène étudié, on pose :

$$Z = \alpha * R + \varepsilon$$

Si R n'est pas liée linéairement au phénomène étudié, on aura intérêt à la discrétiser en plusieurs modalités représentés par des variables indicatrices : R1 R2 R3Rh. et on étudie l'équation :

$$Z = \alpha_1 * R_1 + \alpha_2 * R_2 + \dots + \alpha_h * R_h + \varepsilon$$

I.2. Estimation des paramètres du modèle

On considère un échantillon de N individus indicés $i = 1, \dots, N$. Pour chaque individu, on observe si un certain événement s'est réalisé et l'on note y_i la variable codée associée à l'événement.

On pose
 $\forall i \in [1; N] :$

$$y_i = \begin{cases} 0 & \text{si l'événement s'est réalisé pour l'individu } i \text{ avec probabilité } (1 - p_i) \\ 1 & \text{si l'événement ne s'est pas réalisé pour l'individu } i \text{ avec probabilité } p_i \end{cases}$$

Et $X = (X_1, \dots, X_J)$ le vecteur des J variables explicatives ou prédicateurs de la variable Y

Pour l'estimation des paramètres d'un modèle de régression logistique, la méthode généralement utilisée est celle du maximum de vraisemblance.

Dans le cadre du modèle logit cette fonction s'écrit :

$$L = \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{1-y_i}$$

Puisque p_i dépend du vecteur de paramètre β et du vecteur de variable X_i , on va chercher à maximiser la vraisemblance en les manipulant. Naturellement on ne pourra pas toucher au vecteur X_i , car ce sont les données. Par contre on cherchera quelles sont les valeurs de β qui maximisent cette quantité ℓ .

D'un point de vue pratique, il est plus commode d'utiliser la log-vraisemblance, notée L, celle-ci transformant les produits en somme. Maximiser cette quantité, revient à maximiser la vraisemblance.

$$\ln L = \sum_{i=1}^N y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)$$

A partir de ce moment le problème devient simple, puisque pour trouver le maximum de cette fonction, il suffit d'égaliser sa dérivée à 0. En pratique on utilise des méthodes numériques telles que l'algorithme de Newton-Raphson, pour trouver les valeurs recherchées.

I.3. Interprétation des paramètres

L'interprétation des valeurs des paramètres ne peut être conduite comme en régression. En effet dans un modèle linéaire, le paramètre a une interprétation simple : c'est la variation de Y qui suit une variation d'une unité de X. Par construction, il y a constance de l'effet. Le modèle Logit, n'est pas linéaire, l'interprétation précédente est valable mais uniquement pour la forme linéarisée.

Si l'on réécrit le modèle sous la forme suivante :

$$\frac{p_i}{1-p_i} = e^Z = e^{\beta X} = e^{\beta_0} e^{\beta_1 X_1} \dots e^{\beta_q X_q}$$

On s'aperçoit que la quantité $\exp(\beta)$ (appelée Odds-ratio) représente le facteur par lequel le rapport $p_i/(1-p_i)$ augmente lorsque X varie d'une unité. Cette quantité est indépendante de X. Ainsi, avec cette expression on obtient une valeur de paramètre qui est plus interprétable que le β dans la forme linéarisée (il est difficile de se représenter un Logit) et qui n'a pas l'inconvénient souligné précédemment dans la forme non-linéaire. La seule information réellement utilisable est le signe des paramètres, indiquant si la variable associée influence la probabilité à la hausse ou à la baisse.

I.4. Significativité du modèle

I.4.1. Test du rapport de vraisemblance

Pour tester l'adéquation d'un modèle, on utilise une approche qui consiste à comparer un modèle sans paramètre autre que la constante, et le modèle avec tous les paramètres que l'on veut tester.

$$H_0 : \beta_1 = \dots = \beta_q = 0 \quad \text{vs} \quad H_1 : \text{il existe } i \text{ tel que } \beta_i \neq 0 \quad \forall i=1 \dots q$$

$$LRT = 2(l(a) - l(\hat{a}))$$

$$\text{Sous } H_0 \quad LRT \sim \chi^2_q$$

Avec :

$l(a)$: log de la vraisemblance du modèle estimé à q variables

$l(\hat{a})$: log de la vraisemblance du modèle avec seulement une constante

Règle de décision :

Au niveau de signification α : on rejette H_0 si $LRT > \chi_{q,1-\alpha}$

I.4.2. Test de Wald

Soit le test suivant : $H_0 : \beta_k = 0$ contre $H_1 : \beta_k \neq 0$

Ainsi la statistique de Wald s'écrit :

$$W = \frac{\hat{\beta}_k}{s(\hat{\beta}_k)} \text{ avec } s(\hat{\beta}_k) : \text{l'écart type estimé de } \hat{\beta}_k$$

Sous H_0 : W suit approximativement la loi normale $N(0,1)$.

Règle de décision :

Au niveau de signification α : on rejette H_0 si $W > \phi(1-\alpha)$ avec ϕ : la fonction de réciproque de la loi normale $N(0,1)$.

I.4.3 Autres tests de validité générale du modèle

Des tests de bonne adéquation du modèle ont été proposés par plusieurs auteurs et intégrés dans les logiciels.

- Le critère d'Akaiké

$$AIC = -2 \log L + 2 k$$

Où L est la vraisemblance et k le nombre de paramètres à estimer.

- Le critère de Schwartz

$$SC = -2 \log L + k \log n$$

Où de plus n est le nombre total d'observations.

Ces deux critères sont utiles pour comparer deux modèles différents portant sur les mêmes données. Le modèle pour lequel ces statistiques sont faibles sera préféré

Chapitre 3 : Les techniques de validation du modèle

La construction de l'outil de décision (fonction score) se fait toujours sur un échantillon d'apprentissage ou échantillon de base. Dans ce chapitre, la question est posée de savoir si les performances de l'outil, constatées sur l'échantillon de base, seraient conservées s'il était appliqué à la population toute entière à laquelle il est destiné, encore appelée population cible.

L'analyse discriminante décisionnelle conduit généralement à définir une règle de décision. La première mesure de performance de l'outil discrimination est donnée par les taux de bons classements reposant sur cette règle de décision. C'est un instrument essentiel de sélection du meilleur outil discrimination, mais il n'est pas sans poser quelques problèmes et surtout la liaison entre le seuil de décision et le taux de bons classements. On verra dans ce chapitre d'autres techniques pour examiner cette qualité : courbe de performance du score (courbe de ROC) et les boîtes à moustaches.

I. TAUX DE BONS CLASSEMENTS

La première mesure de performance systématique examinée réside dans les taux de bons classements liés à la règle de décision (encore appelée règle d'affectation). Les taux d'erreurs de classement leur sont bien sur complémentaires.

Le tableau ci-dessous montre la table d'affectation dans le cas de deux groupes à priori notés 1 et 2. Pour chaque individu sa situation réelle (appartenance au groupe 1 ou au groupe 2) est confrontée à son affectation (au groupe 1 ou 2) par la règle de décision liée à l'outil discriminant. L'adéquation de cette affectation avec la réalité est ainsi examinée.

Tableau 26. : Présentation de la table d'affectation

		prévision		
		groupe 0	groupe 1	
réalité	0	a	b	N1
	1	c	d	N2

Chaque individu se trouve dans une et une seule case du tableau. Le dénombrement des observations dans chaque case permet de calculer les taux de bons classements et les taux d'erreurs :

- Pourcentage de bien classés:
 - Du groupe 1 : a/N_1
 - Du groupe 2 : d/N_2

- Du total : $a+d/N_1+N_2$

- Taux d'erreur de classement :
 - Du groupe1 : b/N_1
 - Du groupe2 : c/N_2
 - Du total : $b+c/N_1+N_2$

II. COURBE DE ROC

On rappelle que notre objectif est d'estimer des probabilités des attributs de la variable dépendante dichotomique, prenant les valeurs 1 avec une probabilité p , et 0 avec $1-p$.

On définit la notion de sensibilité comme suit : c'est la probabilité de classer l'individu dans la classe $1 = y$ (test positif) sachant qu'il est effectivement dans cette classe.

Sensibilité = $P(\text{test}+/y=1)$

On définit la spécificité comme étant la probabilité de classer l'individu dans la classe $y=0$ sachant qu'il est effectivement dans cette classe.

Spécificité = $P(\text{test}- / y=0)$.

Courbe de ROC :

La courbe de ROC est utilisée pour tester la capacité du modèle à discriminer les deux groupes

$1 = y$ et $0 = y$

Règle de pouce

Si aire ROC = 0.5 il n'y a pas de discrimination.

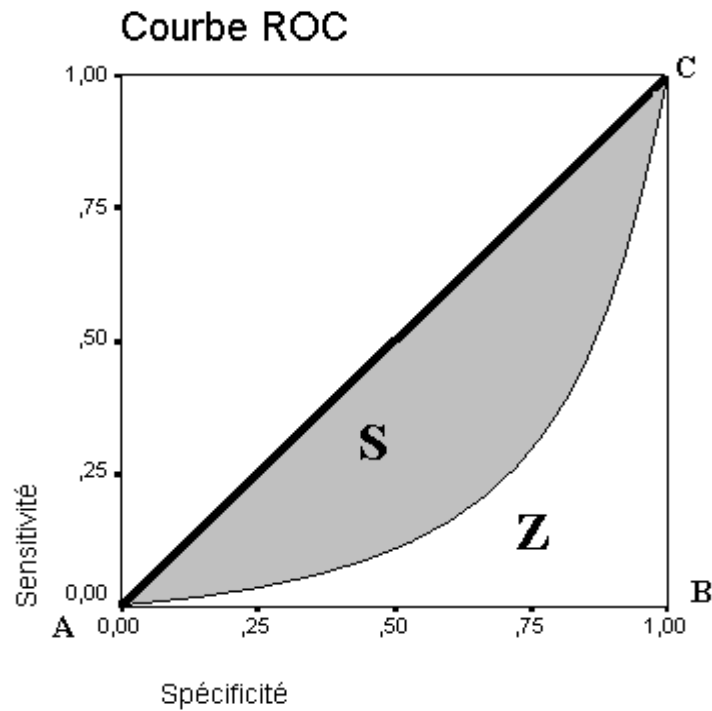
Si aire $0.7 \leq \text{ROC} < 0.8$ la discrimination est acceptable.

Si aire $0.8 \leq \text{ROC} < 0.9$ la discrimination est excellente.

Si aire ROC ≥ 0.9 la discrimination est très excellente.

Un bon modèle doit être bien calibré et permettre une bonne discrimination.

Figure 1. : Présentation d'une courbe de ROC

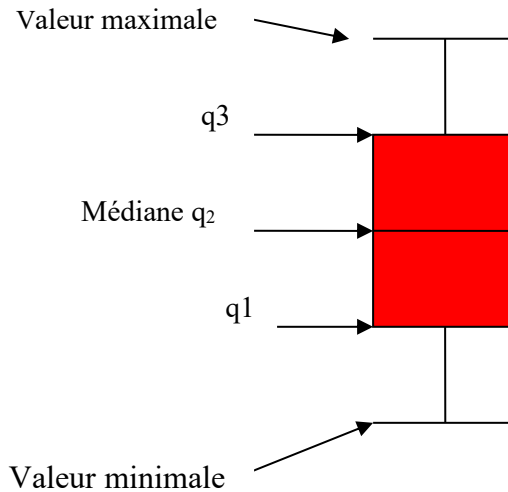


Si cette courbe coïncide avec la diagonale, c'est que le modèle n'est pas plus performant qu'un modèle aléatoire ; alors que, plus cette courbe est proche du coin du carré meilleur est le modèle, car il permet de capturer le plus possible de vrais événements avec le moins possible de faux événements.

III. LES BOITES À MOUSTACHES

La boîte à moustaches (ou « box plot »), inventée par J-W Tukey, est une représentation très populaire, car simple et synthétique, de la dispersion d'une série de données. Elle consiste à placer les valeurs de certains quantiles sur une échelle verticale ou horizontale. Elle permet de visualiser instantanément certaines caractéristiques de tendance centrale, de dispersion et de forme des variables. Elle permet aussi de comparer deux populations.

Figure 2. : Description d'une boîte à moustaches



Les limites supérieure et inférieure des boîtes correspondent aux quartiles supérieur et inférieur. La hauteur de la boîte correspond à l'intervalle interquartile, de sorte que la boîte renferme les 50% du milieu des valeurs d'un groupe. Plus la boîte est grande, plus l'étendue des observations est importante. Les lignes émanant de chaque boîte «les moustaches» vont jusqu'à la plus petite et à la plus grande observation du groupe, à moins d'un intervalle interquartile du bord de la boîte.

Après avoir construit l'outil discriminant, il faut analyser ses performances et notamment sa capacité à discriminer entre les deux groupes. Pour cela on juxtapose sur le même graphique les deux boîtes à moustaches définies respectivement pour les deux groupes et le cas idéal serait que ces deux boîtes à moustaches soient totalement écartées l'une de l'autre dans ce cas on aurait une parfaite discrimination.

APPLICATION:
- Méthode Disqual
- Régression Logistique

Chapitre 1 : Modèle de Scoring du CAM (Disqual)

Dans ce chapitre, nous allons élaborer un modèle de Scoring concernant la clientèle du Crédit agricole du Maroc en se basant sur la méthode DISQUAL, une méthode de discrimination nécessitant l'utilisation de l'ACM appliquée à la base de donnée initiale constituée des variables qualitatives et quantitatives discrétisées, puis on applique l'AFD sur les axes factoriels retenus par l'ACM, cette technique permet de retenir seulement les axes factoriels les plus discriminants, enfin nous allons utiliser la technique de scorecard, qui sert à obtenir les scores associés à chaque modalité. Après la construction du modèle, nous allons déterminer les différentes classes de risque et leurs probabilités de défauts associées qui permettront d'évaluer le client en se basant sur un seuil fixé.

I. APPLICATION DE L'ACM

Nous allons appliquer l'ACM sur la base de données constitué des variables qualitatives, comme une première étape de la méthode Disqual, en utilisant le logiciel XL-stat, nous avons trouvé les résultats présentés dans le tableau ci-dessous :

Tableau7 : Les valeurs propres des axes factoriels

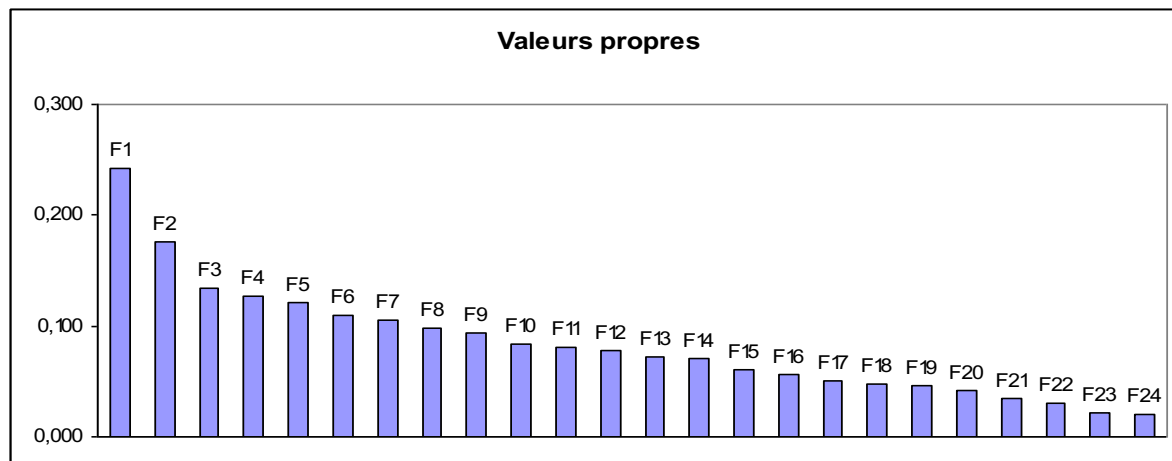
	<i>Valeur propre</i>	<i>Variance %</i>	<i>Cumulé de la variance%</i>
F1	0,242	12,087	12,087
F2	0,176	8,812	20,899
F3	0,135	6,726	27,625
F4	0,127	6,362	33,987
F5	0,121	6,040	40,027
F6	0,110	5,497	45,524
F7	0,105	5,237	50,761
F8	0,098	4,910	55,671
F9	0,094	4,686	60,357
F10	0,084	4,205	64,561
F11	0,080	4,018	68,579
F12	0,077	3,860	72,439
F13	0,071	3,574	76,012
F14	0,070	3,520	79,532
F15	0,061	3,040	82,572
F16	0,056	2,822	85,394
F17	0,051	2,556	87,950
F18	0,048	2,376	90,326
F19	0,046	2,297	92,623
F20	0,042	2,086	94,709
F21	0,035	1,740	96,449
F22	0,030	1,507	97,956
F23	0,021	1,063	99,019
F24	0,020	0,981	100,000

On constate d'après le tableau ci-dessus que le maximum d'inertie expliqué par le premier axe est égal à 12,1%, ainsi que les 24 premiers axes expliquent 100% de l'inertie totale, d'autre part, les 20 premiers axes expliquent 95% de l'inertie totale, c a dire que les taux d'inertie sont trop pessimistes et sous estiment de loin l'information contenue dans les

différents axes factoriels, ce qui paraît normal quand il s'agit de l'étude des modalités ou l'étude en ACM.

Par la suite, nous allons retenir tous les axes puisque l'apport de chaque axe est significatif, comme le montre bien le graphe ci-dessous donné par le logiciel XL-stat:

Graphique1 : Graphe de valeurs propres des axes factoriels



II. APPLICATION DE L'AFD

Nous allons appliquer l'analyse factorielle discriminante sur la nouvelle base de donnée qui est constituée de nouvelles variables qui représentent les coordonnées des anciens individus sur les axes factoriels (24 axes factoriels retenus pour l'étude).

L'application de AFD nous donne le tableau ci-dessous, qui représente les axes factoriels les discriminants, obtenus par la méthode pas à pas (critère lambda de wilks) en SPSS:

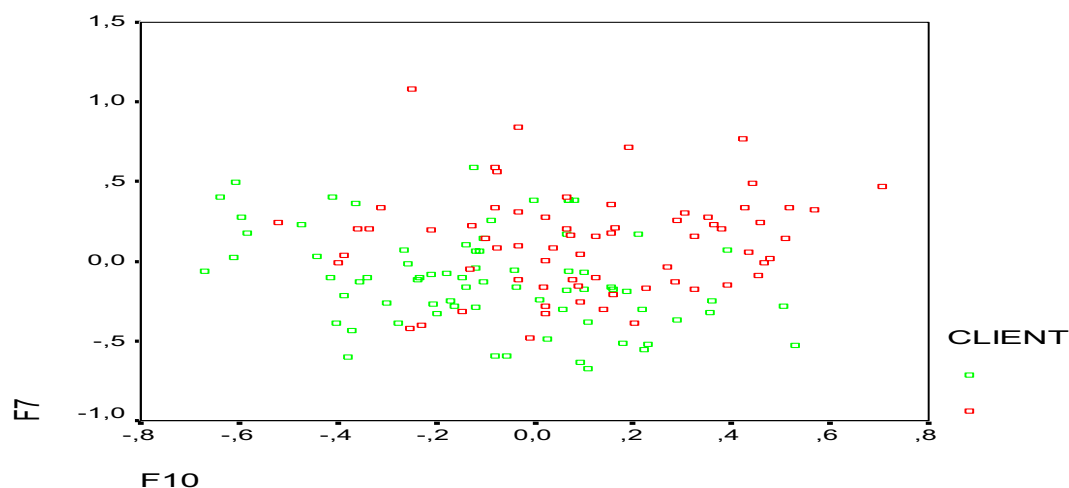
Tableau8 : Les axes factoriels les plus discriminants

AXES	lamda wilks	ddl2	ddl3	Signification
F10	0,87335585	1	139	1,486E-05
F7	0,74866592	1	139	2,1178E-09
F5	0,67937538	1	139	1,7035E-11
F13	0,6106712	1	139	6,5947E-14
F3	0,55450199	1	139	0
F18	0,52755232	1	139	0
F4	0,50326672	1	139	0
F21	0,48094678	1	139	7,9689E-18
F20	0,46280306	1	139	3,1708E-18

On remarque d'après le tableau ci-dessus, que les axes les plus discriminants selon le critère de lambda de wilks sont : F10, F7, F5, F13, F3, F18, F4, F21 et F20 qui sont ordonnés

d'une manière décroissante en terme de leurs pouvoirs de discrimination, comme on constate que les axes F10 et F7 ont des significations les plus grandes, ce qui traduit par conséquent leurs pouvoir de discrimination, et cela est bien vu dans le graphe ci-dessous qui représente la dispersion des clients selon les deux premiers axes, et qui permet bien d'évaluer le pouvoir discriminant des deux axes :

Graphique2 : Graphe de dispersion des clients selon F10 et F7



On remarque d'après le graphe ci-dessus que les deux axes factoriels F10 et F7 discriminent moyennement les deux groupes de clients, alors que les autres axes factoriels, leurs pouvoirs de discriminations sont faibles comme le montre la figure8 annexe II.

Les coefficients de discrimination associés aux axes les plus discriminants, cités ci-dessus, sont représentés dans le tableau ci-dessous, donné par SPSS :

Tableau9 : Coefficients des axes les plus discriminants

<i>F3</i>	1,28675087
<i>F4</i>	-0,8699266
<i>F5</i>	-1,5081493
<i>F7</i>	2,17258134
<i>F10</i>	2,44370003
<i>F13</i>	1,95233594
<i>F18</i>	-1,4995868
<i>F20</i>	1,31323397
<i>F21</i>	-1,5947691

Et par la suite la fonction de discrimination s'écrit alors :

$$(1) \quad Z = -2,5 \cdot 10^{-16} + 1,28 \cdot F3 - 0,86 \cdot F4 - 1,50 \cdot F5 + 2,17 \cdot F7 + 2,44 \cdot F10 + 1,95 \cdot F13 - 1,49 \cdot F18 + 1,31 \cdot F20 - 1,59 \cdot F21$$

La qualité de ce score est justifiée par le coefficient de corrélation canonique concernant cette fonction de discrimination, qui est égale à 0,73 (voir le tableau 31 annexe II), qui montre aussi que la fonction de discrimination permet de récupérer 100% de l'information initiale, c'est-à-dire la totalité de la variance intergroupe est expliquée par cette fonction, ce qui assure son pouvoir discriminant, comme on constate d'après la fonction de score (1) retenu qu'il y a des axes qui ont des coefficients de discriminations négatifs, ce qui diminue par conséquence le score correspondant, c'est à dire augmenter la probabilité d'être un client défaut, comme il y a des coefficients positifs, qui augmentent la probabilité d'être un bon client.

D'autre part, le tableau ci-dessus, qui montre que le modèle composé des axes factoriels les plus discriminants F21, F20, F18, F13, F10, F3, F4, F7 et F5, permet d'atteindre un niveau de discrimination satisfaisant, grâce au niveau de signification du test basé sur Lambda de Wilks qui est assez faible, ce qui permet le rejet de l'hypothèse nulle d'égalité des moyennes de scores des deux groupes de clients.

Tableau10 : Significativité du modèle

<i>Test de la ou des fonctions</i>	<i>Lambda de Wilks</i>	<i>Khi-deux</i>	<i>ddl</i>	<i>Signification</i>
1	0,46280306	103,626019	9	2,9009E-18

En disposant maintenant des scores des axes factoriels, on pourra ainsi déduire les scores des modalités, puisque les axes factoriels sont des combinaisons linéaires des variables initiales, plutôt des modalités, et cela par la méthode Grille de score « **scorecard** ».

(Voire annexe I).

III. LES SCORES ASSOCIES À CHAQUE MODALITE

En appliquant la méthode de **scorecard**, on obtient le tableau ci-dessus, qui représente les scores associés à chaque modalité de la variable explicative initiale, déduite de ceux correspondants aux axes factoriels les plus discriminants :

Tableau11 : les scores des modalités

<i>modalités</i>	<i>Score</i>
sol_1	-6,03380628408208
sol_2	-2,23901198525948
sol_3	6,57167634419071
rentexp_1	-0,594809860630817
rentexp_2	-0,162566495363816
rentexp_3	1,98118091938286
rentfin_1	-0,506403948448118
rentfin_2	0,803620469725308
rentfin_3	-0,427537445728398
serdett_1	0,502604573534479
serdett_2	-0,933774500210447
serdett_3	1,03641689978791

	frng_1	-0,0370079991591307
Solvabilité	frng_2	-0,696628634295106
	frng_3	0,476319352253921
	chargper_1	-0,842081203616441
Rentabilité d'exploitation	chargper_2	0,20902697267556
	chargper_3	1,66080079840045
	ancint_1	-0,484677140734036
Rentabilité financière	ancint_2	1,19888716813112
	ancint_3	-0,414289509982046
	qualsec_1	0,350388374801297
Service de la dette	qualsec_2	0,383616453489741
	qualsec_3	-0,665618428236654
	interet_1	-1,06553067120051
Fonds de roulement/CA	interet_2	1,85474645793279
	interet_3	0,0201281221838403
	partfin_1	0,257226347133493
Charges de personnels	partfin_2	0,603115268894055
	partfin_3	-0,636461867091469
	montech_1	1,44737419083303
ancienneté	montech_2	0,528613610551457
	montech_3	-0,618346490643074
	ratios12_1	0,914722538765002
qualité de secteur	ratios12_2	-1,22252691460682
	ratios12_3	-0,681090334674083
taux d'intérêt		
part financée par CAM		
montant échéance		
montant demandé/Dette		

D'après le tableau ci-dessus ainsi que le tableau qui représente la répartition des clients bons et défauts (Tableau3), on observe que le score des modalités caractérisant le BON client est de signe positif, alors que le DEFAUT client porte un score négatif, donc on peut déduire par la suite les caractéristiques des types de clients, comme il est montré dans le tableau ci-dessus :

Tableau12 : Caractéristiques des deux catégories de clients

<i>BON</i>	<i>DEFAUT</i>
solvabilité supérieure à 0,20	Solvabilité inférieure à 0,20
rentabilité d'exploitation supérieure à 0,12	Rentabilité d'exploitation inférieure à 0,12
Service de la dette inférieur à 50	Service de la dette supérieur à 50
(FRNG/CA)* 360 supérieures à 50 jours	(FRNG/CA)* 360 inférieures à 50 jours
ancienneté supérieure à 3 ans	Ancienneté inférieure à 3 ans
Qualité de secteur stable	Qualité de secteur déclin
montant échéance inférieur à 28000dh / mois	Montant échéance supérieur à 28000
Montant accordé/dette inférieur à 100	Montant accordé/dette supérieur à 100

Le calcul de score pour un client sera la somme des scores (coefficients) des modalités qui le caractérisent ; ainsi on retient la formule suivante :

$$Z(e_k) = \sum \alpha_{ih} * I_{ih}(e_k)$$

Avec :

e_k : le client ou l'entreprise k

α_{ih} : le score de la modalité h concernant la question i, avec $i=1,2,\dots,12$ et $h=1,2,3$

I_{ih} : le variable indicatrice caractérisant la modalité h concernant la question i

IV. Validation du modèle de scoring sur l'échantillon d'apprentissage

La construction d'un modèle de discrimination à partir d'un échantillon d'apprentissage demeure insuffisante, c'est pourquoi il faudrait analyser ses performances et notamment sa capacité à discriminer les deux groupes à travers son test sur des échantillons de test et vérifier s'il reste aussi efficace qu'avec l'échantillon de base.

La taille de l'échantillon dont on dispose est insuffisante pour construire un échantillon de test, c'est pour cette raison que l'évaluation sera basée sur la visualisation du comportement des deux groupes par rapport au score construit ainsi que sur les taux de bon classement qui représentent la première mesure de performance de l'outil discriminant. D'où on abordera la courbe de ROC et les boîtes à moustaches pour faire le test de validation.

IV.1. le taux de bons classements

Pour valider notre modèle de notation calculé par la méthode Disqual, nous avons calculé la matrice de confusion correspondant au modèle à l'aide du logiciel SPSS, et qui est présentée dans le tableau ci-dessous :

Tableau13 : Matrice de confusion (échantillon d'apprentissage)

	<i>CLIENT</i>	<i>0</i>	<i>1</i>	<i>total</i>
<i>Réalité</i>	<i>0</i>	52	16	68
	<i>1</i>	4	69	73
<i>%</i>	<i>0</i>	76,4705882	23,5294118	100
	<i>1</i>	5,47945205	94,5205479	100

Selon la règle d'affectation retenue par le modèle (Tableau30 annexe II), on remarque selon le tableau ci-dessus que le pourcentage de bons clients qui sont bien classés est de

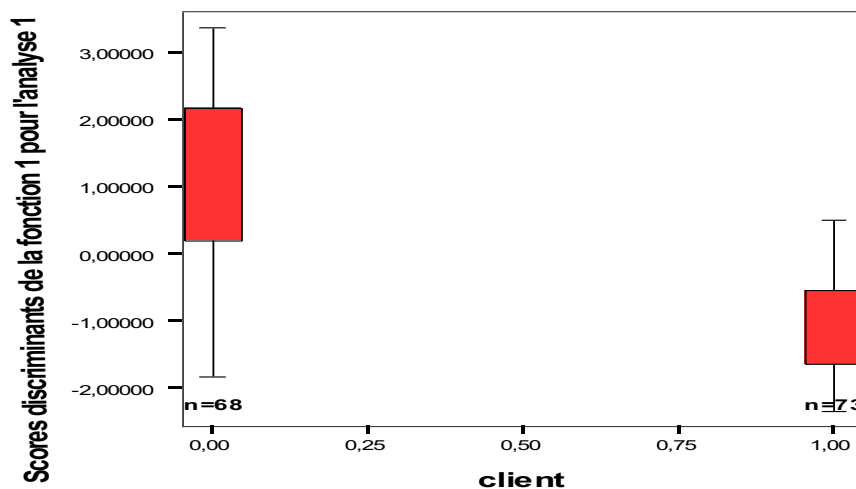
76,5%, alors que le pourcentage concernant les défauts clients est de 94,5%, ce qui signifie que 85,5% des observations sont bien classées, c'est à dire un taux d'erreur apparent est de 14,5%, dont 5,5% de type I (les clients jugés bons par le modèle alors qu'ils sont en réalité défauts) et 23,5% de type II (les clients jugés défauts alors ils sont bons), ce qui est vraiment acceptable.

IV.2. Boîte à moustaches

Comme une deuxième technique de validation du modèle de scoring, nous allons tracer les boîtes à moustaches pour les deux groupes de clients en fonction des scores calculés par le modèle, comme le montre la figure ci-dessous :

Graphique3 : Boîte à moustaches pour les groupes de clients

Boite à moustaches pour chaque catégorie de client

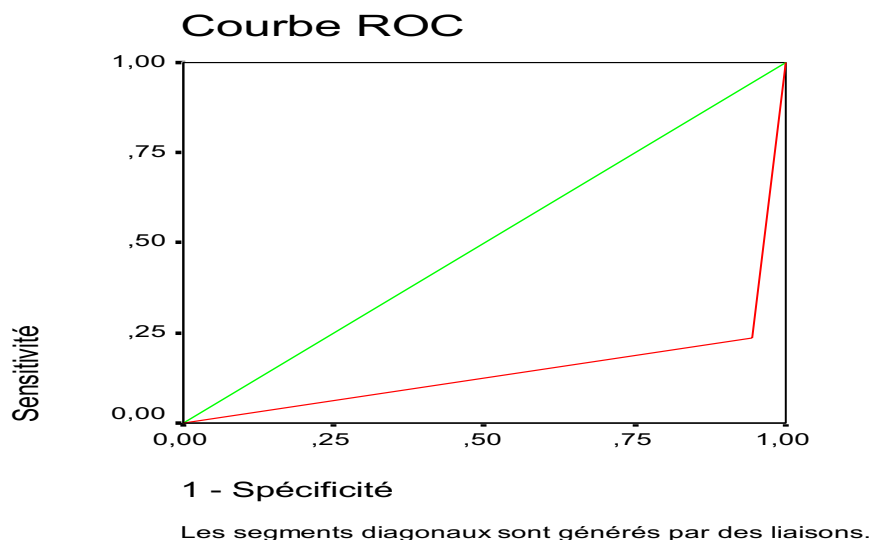


On constate d'après le graphe ci-dessus, qui représente la boîte à moustaches pour les clients bons (0) et défauts (1), que les deux catégories sont bien séparées et bien écartées, et cela est dû au fait que les moyennes des scores des deux groupes sont différentes d'une part, et d'autre part les valeurs correspondantes aux 1er et 3ème quartiles sont bien distinguées, ce qui valide notre modèle de discrimination.

IV.3. la courbe de ROC

Concernant la dernière technique de validation correspondante à la courbe de ROC et qui permet de mesurer le degré de détection des vrais événements à travers un indice appelé indice de Gini calculé à l'aide du sous-surface (Z), comme le montre bien la figure ci-dessous :

Graphique4 : Courbe de ROC (méthode Disqual)



On constate d'après la figure ci-dessus que la courbe est éloignée de la droite du diagonal et proche du coin du carré, c'est à dire que la surface sous courbe est petite, et cela veut dire que notre modèle capture le plus possible de vrais événements que ça soit les clients bons ou défauts avec évidemment le moins possible de faux événements, et ceci est encore justifié par l'indice de Gini qui est égale à $G = 1 - 2Z = 1 - 2 \times 0,145 = 0,71$

V. CLASSE DE RISQUE ET SEUIL D'AFFACTATION

V.1. classe de risque

Après l'élaboration du modèle de discrimination et le calcul des scores pour les clients et leurs probabilités de défaut a posteriori, nous allons définir par la suite les classes de risque et les probabilités de défauts associés.

L'objectif fixé dès le début de notre projet, c'était de construire 8 classes homogènes de risques, dont une représente la zone très risquée là où on va rejeter directement la possibilité d'octroi de crédit au client qui y appartient, du fait qu'elle sera caractérisée par un score négatif, et les autres classes d'acceptation seront significativement différentes, homogènes et dont les probabilités de défauts sont faibles.

En appliquant la méthode de classification hiérarchique (CH) (annexe I) sur les scores correspondants à l'ensemble des individus obtenus par l'analyse factorielle discriminante, on obtient les classes de scores et leurs probabilités moyenne de défaut, comme il est présenté dans le tableau ci-dessous :

Tableau14 : Classe de risque et la probabilité de défaut associé

Classe de risque	Intervalle de scores	Probabilité de défaut (PD)
1	$[\infty ; 0 [$	-
2	$[0 ; 0,3[$	0,0283687
3	$[0,3 ; 0,7[$	0,0283687
4	$[0,7 ; 1,4[$	0
5	$[1,4 ; 1,8[$	0
6	$[1,8 ; 2,5[$	0
7	$[2,5 ; 3[$	0
8	$[3 ; \infty [$	0

D'après le tableau ci-dessus, la classe 1 est la classe où nous allons rejeter la demande d'octroi de crédit qui est caractérisée par des scores négatifs, alors que les classes 2 et 3 représentent les classes où nous allons accepter la demande d'octroi de crédit mais la probabilité que le client d'être un client défaut est de 0,01418, tandis que les autres classes de risques ont des probabilités de défauts nulles, ce qui nous pousse par la suite de nous contenter seulement sur trois types de classes de risque homogènes et significativement différentes, comme il est présenté dans le tableau ci-dessus :

Tableau15 : Les classes de risques homogènes et PD

Classe de risque	Probabilité de défaut
$[\infty ; 0 [$	-
$[0 ; 0,7[$	0,05436
$[0,7 ; \infty [$	0

V.2. Le seuil d'affectation

Tableau16 : Les stratégies d'affectations des clients

D'après le tableau ci-dessus, qui montre les différentes erreurs associées aux différents

	Politique n°1	Politique n°2	Politique n°3	Politique n°4	Politique n°5
<i>Seuil de sélection</i>	Note > -0,5	Note > 0	Note > 0,5	Note > 1	Note > 1,25
<i>Nombre de dossier</i>	141	141	141	141	141
<i>Taux de bon classement</i>	80,5%	85,5%	85,1%	79,5%	77,5%
<i>Taux d'erreur</i>	19,5%	14,5%	14,9%	20,5%	22,5%
<i>Erreur de type I</i>	22%	5,5%	5%	4,5%	3%
<i>Erreur de type II</i>	17%	23,5%	24,8%	36,5%	42%

stratégies ou politiques d'octroi de crédit selon le seuil choisi, nous constatons que le seuil qui assure un minimum d'erreur (taux d'erreur) est égal à 0 qui représente un taux de bon classement pour les bons et les défauts clients et qui est de 85,5%, c'est à dire un taux d'erreur égal à 14,5%, Comme nous remarquons aussi que l'erreur de type I est diminué avec

l'augmentation du seuil, alors l'erreur de type II augmente, par la suite, on va fixer le seuil de zéro pour la règle de décision, comme il est présenté dans le tableau ci-dessous :

Tableau17 : La règle de décision

SEUIL	Inférieur à 0	Supérieur à 0
Décision	De ne pas faire de crédit	Faire de crédit

Conclusion

Après l'élaboration du modèle de Notation basé sur la méthode Disqual qui nous a permis de déterminer les scores associés à chaque modalité, de construire des classes de risques ainsi que de calculer les probabilités de défauts associées, et qui a été jugé bon à travers plusieurs tests à savoir : le taux de bons classements, courbe de ROC et les boites à moustaches, nous aurons recours dans le chapitre suivant à la construction d'un modèle de Scoring en procédant par la Régression Logistique.

Chapitre 2 : Modèle de Scoring du CAM (Régression Logistique)

Dans ce deuxième chapitre, nous allons essayer de construire un modèle de notation qui tient en compte seulement des modalités qui étaient retenues par la méthode de sélection de variables (méthode descendante : rapport de vraisemblance), pour estimer la probabilité a posteriori qu'un client ne règle pas ces engagements avec la banque. Ensuite, on va déterminer les classes de risques, les probabilités de défauts moyennes et le seuil d'affectation des clients.

I. PRESENTATION DE LA BASE DE DONNEES

Notre base de données que nous allons utiliser pour construire le modèle logistique est issue de celle utilisée pour l'élaboration du modèle scoring par la méthode Disqual (voir Tableau27 annexe), mais la seule différence réside dans les variables explicatives utilisées dans le modèle logistique qui représentent les variables indicatrices associées aux modalités des variables qualitatives et quantitatives discrétisées après l'élimination des modalités les plus courantes pour chaque variable, comme le montre bien le tableau ci-dessous :

Tableau18 : Les modalités supprimées (références)

<i>Variable explicative</i>	<i>Poids des modalités%</i>			<i>Modalité supprimée</i>
	1	2	3	
Solvabilité	8,5	62	29,5	2
Rentabilité d'exploitation	46	36	18	1
Rentabilité financière	44	37	19	1
Service de la dette	39	41	20	2
FRNG/CA	42	23	35	1
Charges de personnels	53	23	24	1
Ancienneté	30	27	43	3
Intérêt	39	22	39	3
Montant échéance	11	35	54	3
Qualité du secteur	34,8	29,8	35,4	3
Part financier CAM	26	34	40	3
Ratios 12	48	19	32	1

Par conséquent, la nouvelle base de données est constituée de 24 variables indicatrices (explicatives), et 141 observations.

II. CONSTRUCTION DU MODELE LOGISTIQUE

II.1. Sélection des variables du modèle

Selon la méthode descendante (Rapport de vraisemblance) usée par le logiciel SPSS, nous allons retenir les variables suivantes qui sont jugées significatives et qui sont estimées par la méthode du maximum de vraisemblance après 16 itérations, comme le montre bien le tableau ci-dessous :

Tableau19 : Les coefficients des variables retenus

Variable	Coefficient
SOL1	4,15272155
SOL3	-1,81666161
RENTEXP3	-1,36832349
RENTFI2	-1,63050026
SERDE1	-2,70012535
FRNG3	-1,34160921
QUALSEC2	-0,9800191
Constante	2,67882692

Nous constatons d'après le tableau ci-dessus, que les variables retenues pour la construction du modèle logistique (variables pertinents) ainsi que leurs coefficients associées sont soit de signe positif, qui augmente par conséquence la probabilité d'être un défaut client, ou de signe négatif, ce qui augmente la probabilité d'être un bon client.

II.2. Significativité du modèle

II.2.1. test de Wald

$$H_0: \beta_i = 0 \quad \text{VS} \quad H_1: \beta_i \neq 0 \quad i=1 \dots 7 \quad \alpha = 5\%$$

Règle de Décision :

$$\text{On rejette } H_0 \text{ si } W > \phi(1 - \alpha) = \phi(0.95) = 1.65$$

Donc nous avons calculé la statistique de Wald (W) pour les différentes variables retenues afin de juger leurs niveaux de signification, comme il est présenté dans le tableau ci-dessous :

Tableau20 : La statistique de Wald pour les variables retenus

Variabes	Wald (W)
SOL1	9,60199497
SOL3	7,30475225
RENTEXP3	3,05629828
RENTFI2	6,66703612
SERDE1	16,4020329
FRNG3	6,89878988
QUALSEC2	4,48996317
Constante	16,1796654

Donc on remarque d'après le tableau ci-dessus que toutes les variables sont significatives différentes de zéro au seuil de 5% puisque toutes les valeurs de la statistique de Wald (W) sont supérieures à 1,65.

II.2.2. test du Rapport de vraisemblance

Nous allons tester la significativité du modèle constitué de toutes les variables pertinentes avec celle correspondante au modèle réduit, à l'aide de la statistique LRT calculé par le logiciel SPSS.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \quad \text{vs} \quad H_0 : \text{il existe } i \text{ tel que } \beta_i \neq 0 \quad \forall i = 1 \dots 7$$

On d'abord :

$$2l(a) = 2 \log \text{vraisemblance (complet)} = -150,158$$

$$2l(\hat{a}) = 2 \log \text{vraisemblance (réduit)} = -350,345$$

$$\text{Donc LRT} = 2l(a) - 2l(\hat{a}) = 200,187 > \chi_{7; 0,05}^2 =$$

Conclusion :

On rejette l'hypothèse H_0 au seuil de 5%, ce qui signifie que la différence entre le modèle complet qui contient toutes les 7 variables explicatives et le modèle réduit qui contient seulement la constante est significativement différent de zéro.

II.2.3. Autres tests de validité générale du modèle

- **Le critère d'Akaiké**

$$AIC = -2 \log L + 2k$$

$$\text{Avec } k = 7 \quad \text{et } \log L = l(a) = -150,158 / 2 = -75,08$$

$$\text{Donc on trouve } AIC = 150,158 + 14 = 164,158$$

- **Le critère de Schwartz**

$$SC = -2 \log L + k \log n$$

$$\text{Avec } n = 141$$

$$\text{Donc on a } SC = 150,158 + \log 141 = 155,106$$

Le critère d'Akaiké (AIC)	Le critère de Schwartz (SC)
164,158	155,106

Conclusion

Les faibles valeurs concernant les deux statistiques calculées SC et AIC pour le modèle retenu par la régression logistique justifient bien la validité du modèle.

III. Validation du modèle par l'échantillon de construction

III.1. le taux de bons classements

Après la mise en place d'un outil de discrimination (modèle logistique), Les taux de bon classement sont systématiquement examinés, leur calcul consiste à comparer les affectations effectuées par le modèle avec celles d'origine, après le calcul des probabilités de défaut a posteriori pour toutes les entreprises.

Si Probabilité (p_i) est supérieure à 0,5 alors l'individu i sera affecté au groupe n° : 1 (Défaut)

Si Probabilité (p_i) est inférieure à 0,5 alors l'individu i sera affecté au groupe n° : 0 (Bon)

Le taux de bons classements concernant notre échantillon d'étude est de 71,63%, c'est à dire un taux d'erreur égal à 28,37%, ce qui est bon et valide notre modèle proposé, ce qui est présenté dans le tableau ci-dessous :

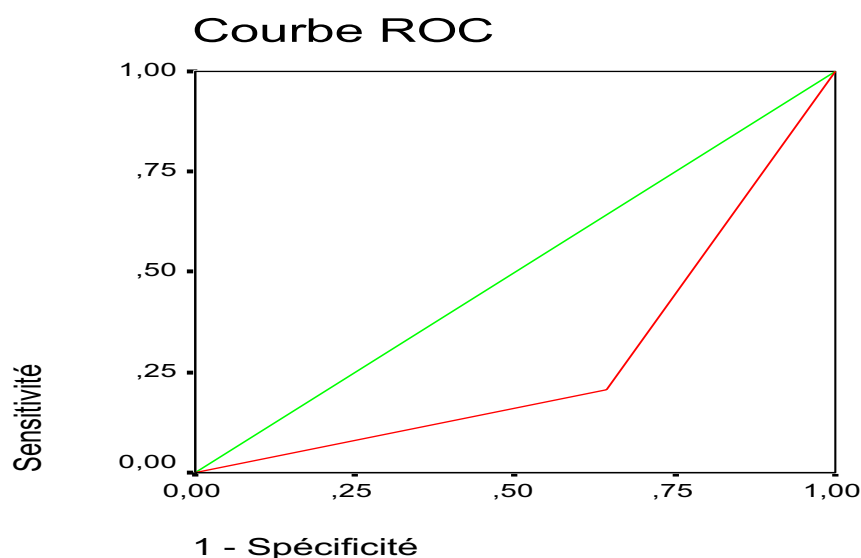
Tableau21 : Taux de bons classements des deux groupes de clients

Taux de bons classements (BON)	Taux de bons classements (Défauts)
39,7%	60,3%
Taux de bons classements global	71,63%

III.2. Courbe de ROC

Comme une deuxième méthode de validation du modèle de discrimination issue de la Régression logistique, nous allons utiliser la courbe de ROC pour évaluer sa performance, comme le montre le graphique ci-dessous :

Graphique5 : Courbe de Roc pour les deux groupes de clients



Les segments diagonaux sont générés par des liaisons.

D'après la figure ci-dessus on remarque que la courbe s'éloigne un peu de la première bissectrice, ce qui implique que notre modèle est moyennement bon.

En plus on a l'indice de Gini = $(1-2Z) = (1-2* 0,281) = 0,43$

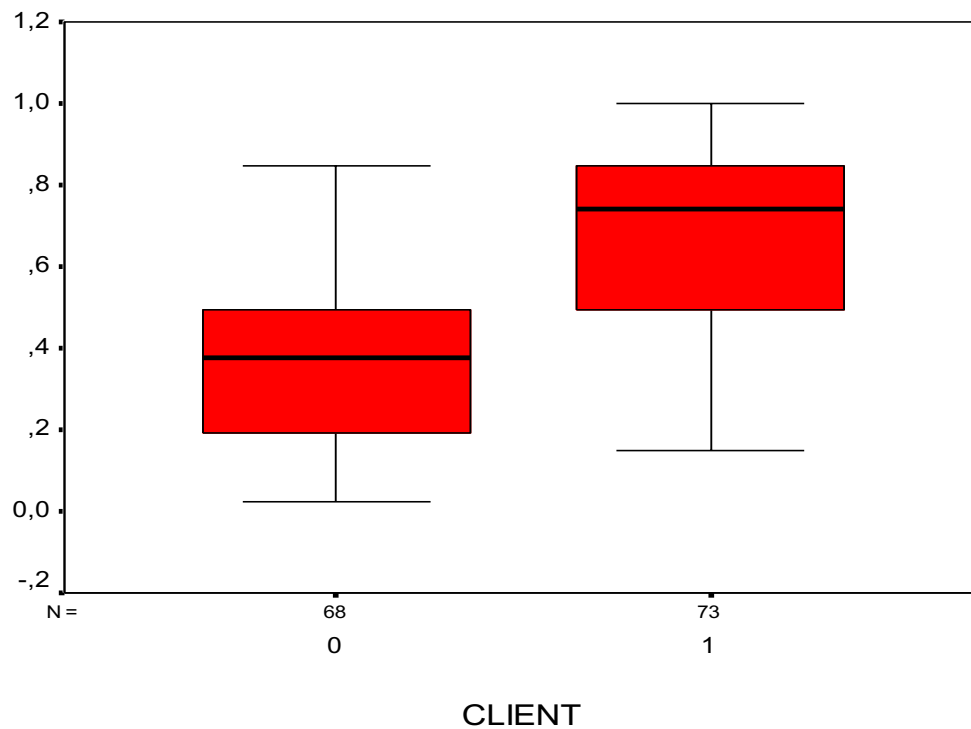
Conclusion :

La valeur de l'indice de Gini est moyenne, et nous indique une fois de plus que notre modèle de scoring par la régression logistique est moyennement bon.

III.3. Boîtes à moustaches

Concernant maintenant la dernière méthode de validation du modèle de notation, et celle correspondante à la boîte à moustaches, comme le montre le graphique ci-dessous :

Graphique6 : Boite à moustaches pour les deux groupes de clients



On constate d'après la figure ci-dessus que les boîtes à moustaches des deux catégories de clients (BON) et (DEFAULT) s'éloignent un petit peu l'une de l'autre, et que les probabilités moyennes estimées pour les deux catégories de clients sont différentes, ce qui signifie que notre modèle est moyennement bon.

Conclusion

Le modèle calculé par la méthode de régression logistique, qui contient les variables explicatives retenues par la méthode Descendante (Rapport de Vraisemblance), est validé par plusieurs tests, en plus il a donné des résultats moyennement satisfaisants en terme de taux de bon classement et taux d'erreur associé.

IV. LES CLASSES D'AFFECTION DES CLIENTS

Nous avons construit des classes de risques, en utilisant la CHA en SPSS, qui nous a fourni aussi les probabilités associés à chaque classe de risque, comme le montre le tableau ci-dessous :

Tableau24 : Classe de risque par la méthode logistique

<i>Classe de risque</i>	<i>Classe de risque pour Pi</i>	<i>Probabilité de défaut (PD)</i>
1	[0 ; 0,08[0
2	[0,08 ; 0,17[0,0125
3	[0,17 ; 0,23[0,0205
4	[0,23 ; 0,35[0,0405
5	[0,35 ; 0,4[0,05
6	[0,4 ; 0,45[0,025
7	[0,45 ; 0,5[0,1625
8] 0,5 ; 1[-

On constate d'après le tableau ci-dessus que la classe7 a une probabilité de défaut égale à 0,1625 ce qui est risqué, alors que les autres classes ont des probabilités de défauts presque égales, c'est pourquoi nous avons réduit le nombre de classe en 7 classes homogènes et significativement différentes, comme le montre bien le tableau ci-dessus :

Tableau : Classes de risques et leurs PD (Logistique)

classes	classes de risque pour Pi	Probabilité de défauts PD
1	[0 ; 0,08[0
2	[0,08 ; 0,17[0,0125
3	[0,17 ; 0,23[0,0205
4	[0,23 ; 0,35[0,0405
5	[0,35 ; 0,45[0,075
6	[0,45 ; 0,5[0,1625
7] 0,5 ; 1[-

V. SIMULATION DE SEUIL D'AFFAECTATION

Tableau25 : Stratégies d'affectation des clients

	<i>Stratégie1</i>	<i>Stratégie2</i>	<i>Stratégie3</i>	<i>Stratégie4</i>
<i>Seuil d'affectation</i>	0,5	0,6	0,45	0,40
<i>Taux bons classements</i>	71,6%	69,5%	71%	70,9%
<i>Taux d'erreur</i>	28,3%	30,5%	29%	29,07%

Donc on déduit du tableau ci-dessus que la meilleure stratégie qui permet d'obtenir un taux de bons classements assez bon est celle correspondante au seuil de 0,5.

Tableau26 : Règle de décision

BON	DEFAUT
$P_i < 0,5$	$P_i > 0,5$

P_i : la probabilité a posteriori estimé par le modèle concernant le client i

Il sera question maintenant de comparer ce modèle retenu par la méthode Logistique avec celui trouvé par la méthode Disqual.

VI. Comparaison entre le modèle logistique et le modèle scoring (Disqual)

VI.1. Taux de bons classements

Nous avons utilisé la notion du taux de bons classements pour pouvoir comparer les deux méthodes de discriminations, Disqual et la Régression logistique, en terme de performance et de pouvoir de discrimination, comme il est présenté dans le tableau ci-dessous :

Tableau22 : le taux de bons classements pour les méthodes de discriminations

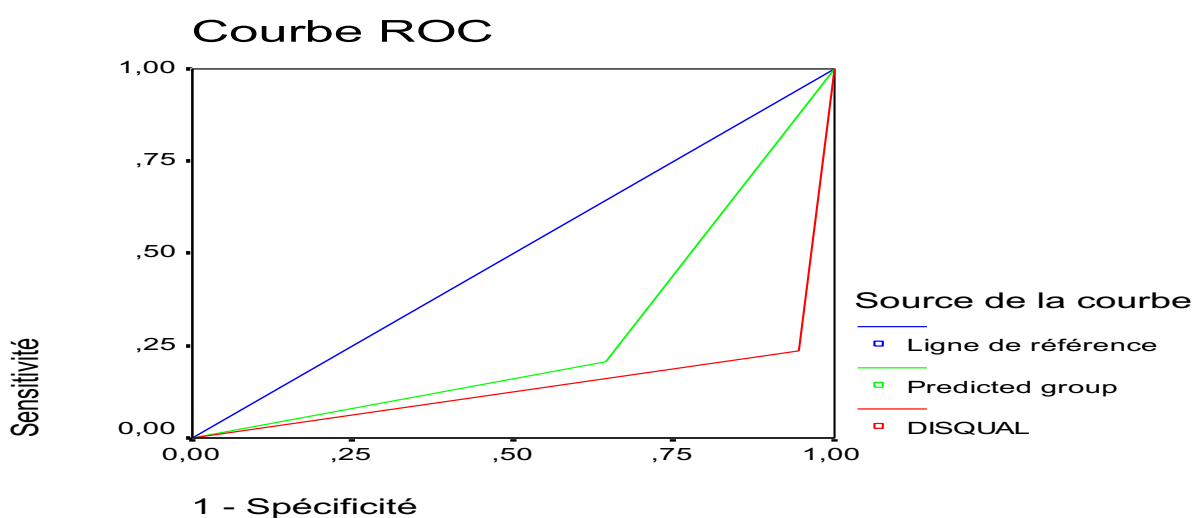
<i>méthode</i>	<i>Taux bons classements (BON)</i>	<i>Taux bons classements (DEFAUT)</i>	<i>Taux bons classements global</i>	<i>Taux d'erreur</i>
LOGISTIQUE	39,70%	60,30%	71%	29%
DISQUAL	76,5%	94,5%	85,5%	14,5%

D'après le tableau ci-dessus, il s'avère bien que le modèle basé sur la méthode Disqual est plus performant en terme de taux de bons classements que celui fondé sur la régression logistique.

VI.2. Courbe de ROC et l'indice de Gini

Comme une deuxième techniques de comparaison entre les deux modèles de discriminations, nous avons utilisé la courbe de ROC, comme le montre le graphique ci-dessous :

Graphique7 : Comparaison entre les deux méthodes de discrimination par la courbe de ROC



On observe selon la figure ci-dessus que la courbe qui représente la méthode de discrimination (Disqual) occupe une surface plus grande que celle occupée par la courbe logistique ce qui signifie que la méthode Disqual est plus pertinente et plus approuvable que celle de la régression logistique, comme le montre aussi le tableau ci-dessous :

Tableau23 : Indice de Gini pour les deux méthodes de discriminations

<i>méthode</i>	<i>Surface occupée (Z)</i>	<i>Indice de Gini (G)</i>
LOGISTIQUE	0,281	43,8%
DISQUAL	0,145	71%

D'après le tableau ci-dessus, on remarque que les valeurs de l'indice de Gini pour les deux méthodes justifient et valident ce qui a été trouvé par le taux de bons classements et aussi les courbes de ROC associées, ce qui veut dire en conclusion que le modèle calculé par la méthode Disqual est préférable au modèle obtenu par la méthode Logistique.

Toutefois, le Scoring reste une technique qui nécessite un suivi régulier, vu le besoin d'actualiser les critères explicatifs, vérifier l'efficacité du score et l'exactitude des prévisions.

Projet fin d'étude
2006-2007

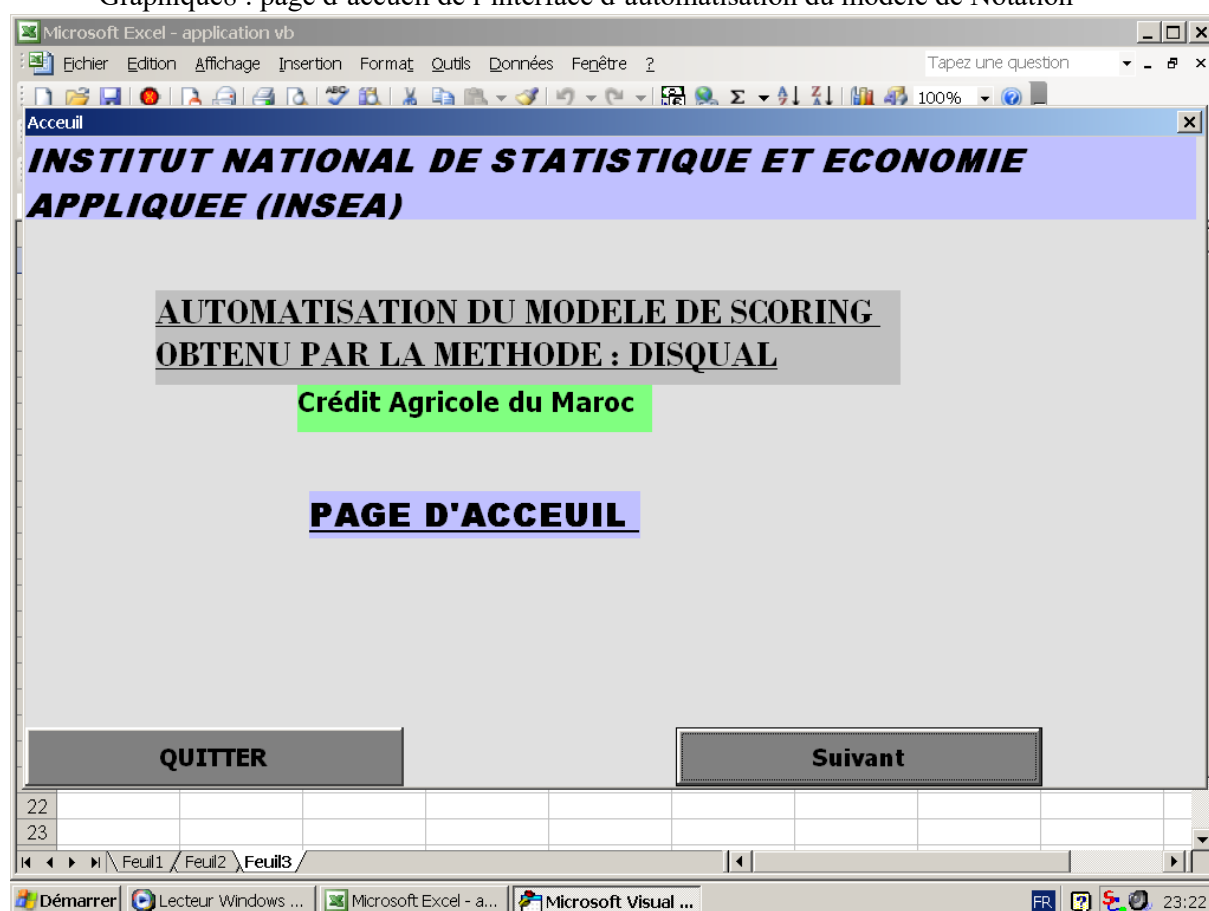
Informatisation Du Modèle de Scoring : DISQUAL

Informatisation du système de Notation Interne

Afin de faciliter la compréhension et l'application du modèle de notation des clients retenu par la méthode de Disqual, on va procéder par l'informatisation de l'enchaînement des scénarios d'élaboration de score pour les clients du Crédit Agricole du Maroc, en utilisant le langage de programmation Visuel basic (Vb-Excel).

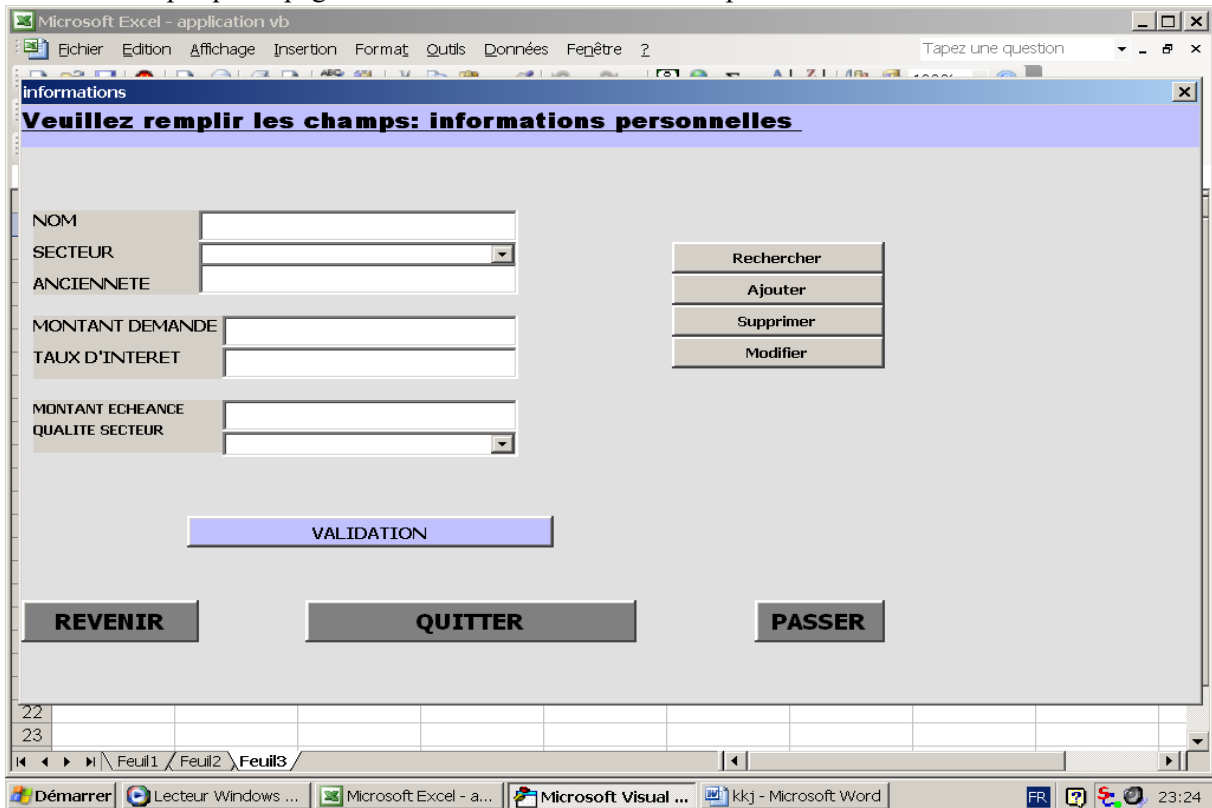
Ainsi, pour attribuer une note à un client, ce dernier doit fournir des informations personnelles qui le concernent directement et celles qui reflètent l'état financière de l'entreprise, ceci nous permettra par la suite de déterminer le score et la probabilité de défaut correspondante.

Graphique8 : page d'accueil de l'interface d'automatisation du modèle de Notation



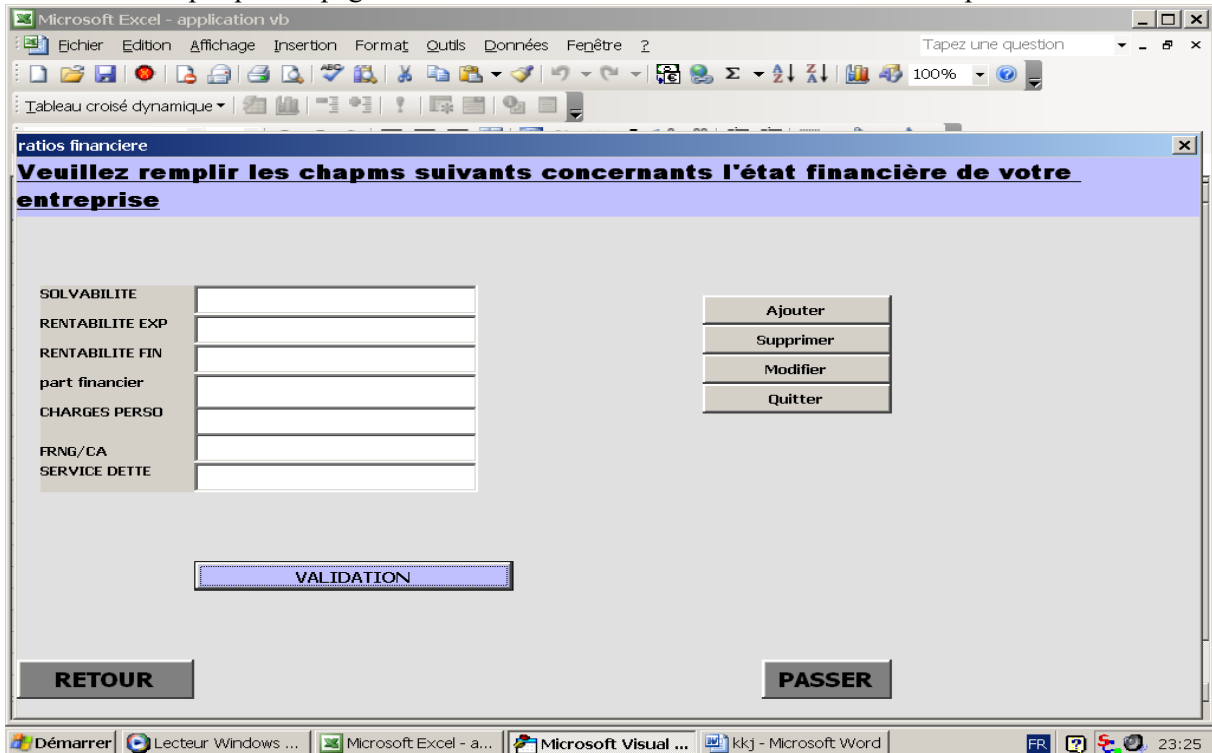
La figure ci-dessus présente la page d'accueil concernant l'interface d'exécution du modèle de scoring obtenu par la méthode Disqual pour le Crédit agricole du Maroc.

Graphique9 : page d'introduction des informations préliminaires du client



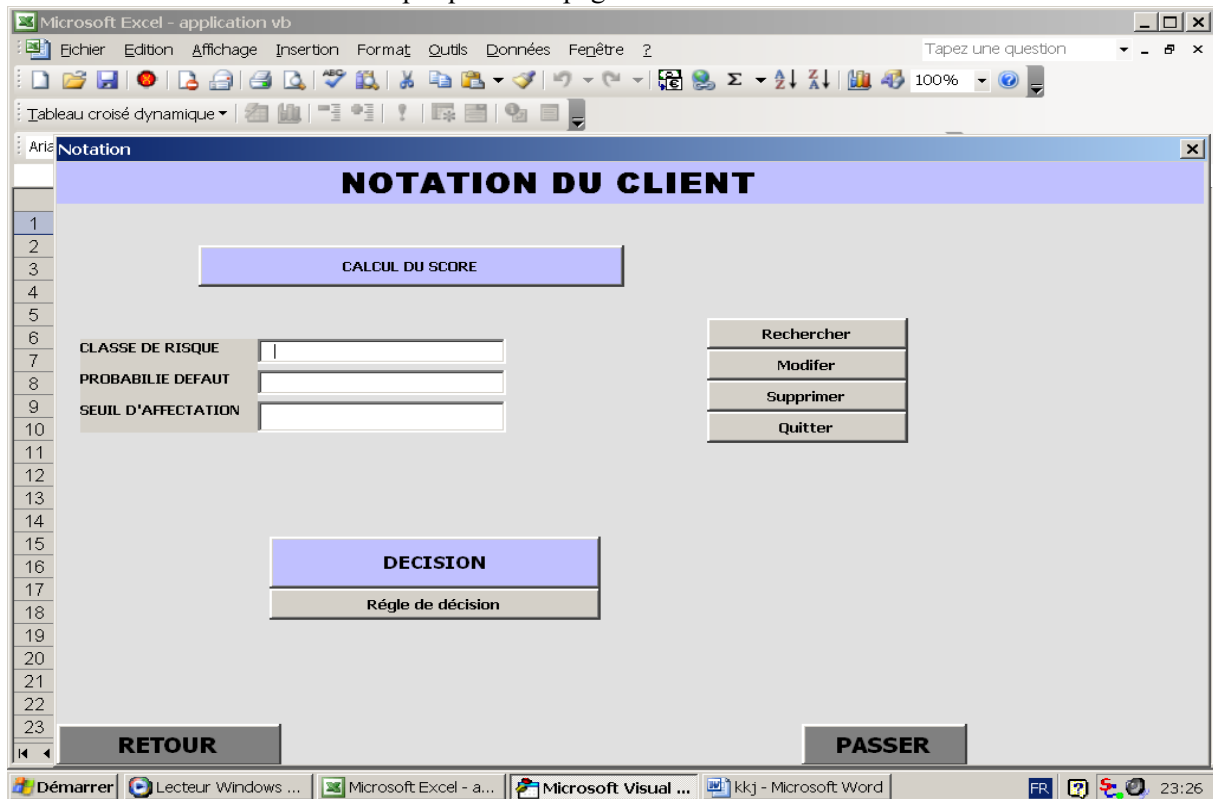
La figure ci-dessus présente les informations préliminaires que doit remplir le client, concernant le nom de la société, le secteur d'activité, la durée d'existence dans le secteur, et le montant demander.

Graphique10 : page d'insertion des informations financières de l'entreprise/client



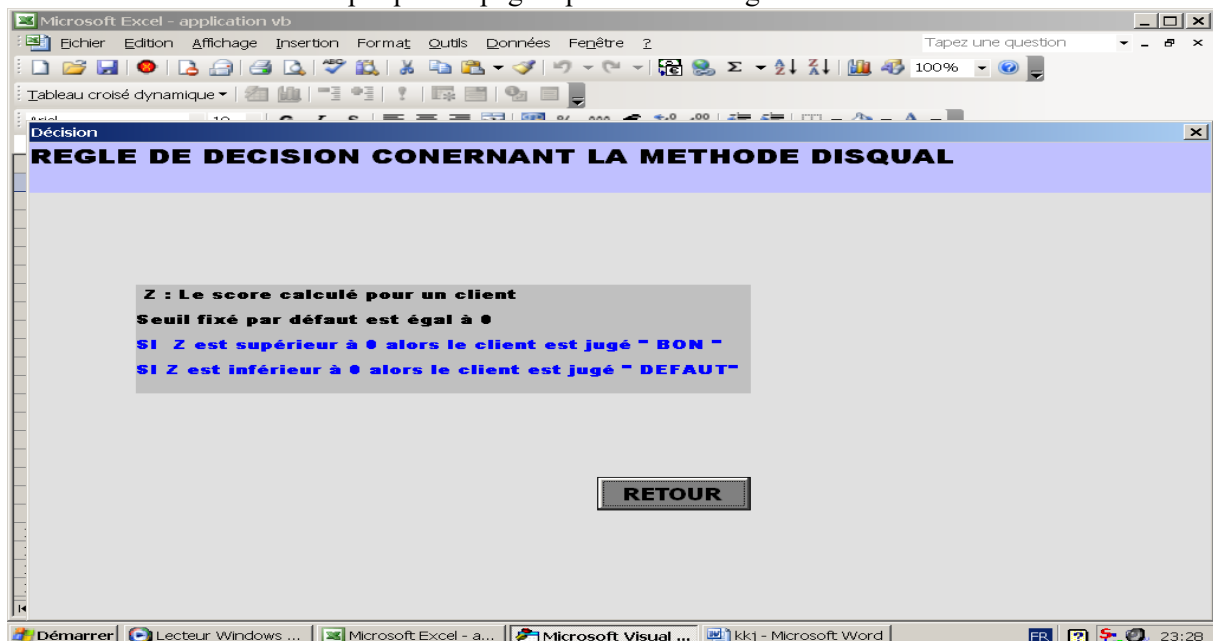
La figure ci-dessus présente les informations financières qui reflètent l'état de l'entreprise, et qui permet d'évaluer le client à savoir : le ratios de solvabilité, le ratios de rentabilité de l'entreprise indiqué par les deux ratios (rentabilité d'exploitation et rentabilité financière), ainsi que des ratios qui donnent une idée sur le poids des dettes.

Graphique11 : la page de notation du client



La figure ci-dessus présente la page qui exhibe le score du client, la classe de risque et la probabilité de défaut moyenne correspondante, ce score permet d'évaluer la qualité du client en se basant sur le seuil fixé.

Graphique12 : page représentant la règle de décision



La figure ci-dessus présente la page d'aide qui donne une idée sur la méthode d'élaboration d'une décision concernant la méthode Disqual, et qui donne par défaut le seuil d'affectation de client égale à zéro.

Conclusion

Au terme de ce travail, nous avons abouti à construire deux modèles de notation de la clientèle du crédit agricole du Maroc avec notamment l'identification des classes de risques et le calcul des probabilités de défaut associées. Le premier modèle fondé sur la méthode Disqual qui permet de déterminer les scores correspondants aux différentes modalités concernant les variables retenues au début ,en se basant dans un premier lieu sur l'analyse correspondance multiple (ACM), et en second lieu sur l'analyse factorielle discriminante (AFD) appliquée sur les axes factorielles obtenus par l'ACM, ensuite on a utilisé la technique scorecard qui nous a permis de trouver les scores relatifs à chaque modalité à travers les scores correspondants aux axes factoriels les plus discriminants .Le deuxième modèle construit à partir de la régression logistique nous a permis d'estimer les probabilités a posteriori de défaut en se basant sur les variables sélectionnées. Le premier modèle nous a permis d'obtenir des taux de bon classement meilleurs que ceux du deuxième modèle, c'est pourquoi il a été retenu comme le modèle le plus pertinent.

Les deux modèles ont mis en évidence le poids important du critère de la solvabilité sur la qualité du client et la capacité de chaque client à régler ces engagements, ladite solvabilité a été retenue comme la variable la plus discriminante qui sépare mieux les deux catégories de clients.

D'autre part, ce projet nous a permis de mettre en œuvre une bonne partie de nos connaissances théoriques acquises durant notre formation à l'Institut, découvrir de près le secteur financier, aborder l'approche statistique appliquée à la théorie financière et de développer l'esprit d'analyse et de synthèse en se confrontant aux problématiques réelles qui se posent au niveau du contrôle et du gestion interne de l'organisme concerné.

Nonobstant cela, nous avons rencontré pas mal de problèmes notamment l'indisponibilités de

certaines données, la qualité de l'information existante ainsi que la contrainte du temps. Ceci a bien affecté la performance et le bon déroulement du projet.

Enfin, la performance de l'approche scoring se voit dans l'identification et la détection des risques de la clientèle, attribuer une note à chaque client et calculer les probabilités de non remboursements de ces engagements, ceci permettra bien d'améliorer la prévision des risques en vue d'une couverture optimale de ces derniers et une optimisation des fonds propres usés pour faire face aux risques encourus. Toutefois, le Scoring présente quelques faiblesses surtout en cas d'une mauvaise qualité de données, en plus il nécessite des données sur un grand nombre de prêts historiques et exige beaucoup de données pour chaque prêt, ce que nous avons déjà confronté durant notre stage au Crédit Agricole du Maroc.

ANNEXE I

Méthode Grille de score (« scorecard »)

Méthode Grille de score c'est une méthode qui permet d'obtenir le score de modalités des variables qualitatives à partir de score des axes factorielles les plus discriminants, obtenu par l'application de l'analyse factorielle discriminante (ACM) au tableau disjonctif complet (TDC) et l'AFD.

On a $S = \sum_{j=1}^k d_j Z^j$, avec $Z^j = XU^j$ Avec U^j : coordonnées des modalités sur l'axe n°j

$$\text{Alors, } S = \sum_{j=1}^k d_j Z^j = \sum_{j=1}^k d_j XU^j = X \underbrace{\sum_{j=1}^k d_j U^j}_{\text{grille de score}}$$

Avec k : le nombre d'axes factorielles retenus

$$\begin{pmatrix} \cdot \\ d_j \\ \cdot \end{pmatrix} = \mathbf{V}^{-1} (\mathbf{g}_1 - \mathbf{g}_2) = \begin{pmatrix} \cdot \\ \frac{\mathbf{z}'_1 - \mathbf{z}'_2}{V(\mathbf{z}^j)} \\ \cdot \end{pmatrix}$$

X : le tableau disjonctif complet (TDC)

d_j : le score associé à j'ème axe factorielle

Classification automatique

Les techniques de classification automatique permettent d'obtenir d'une manière automatique des groupes d'individus en se basant sur leurs descripteurs, ces groupes sont censé être homogènes de point de vue les descripteurs étudiés. Les techniques de classification font appel à des algorithmes numériques et non pas aux calculs algébriques.

La classification ascendante hiérarchique, est une des techniques les mieux adaptées aux données quantitatives, elle se base sur un critère de groupement, qui se rapporte soit à la distance entre groupes ; en considérant l'existence d'une distance : distance moyenne entre groupes $d(X, Y) = \text{moyenne}_{\substack{x \in X \\ y \in Y}} d(x, y)$, distance du saut minimal $d(X, Y) = \inf_{\substack{x \in X \\ y \in Y}} d(x, y)$,

distance entre centres de gravité $d(X, Y) = d(G_x, G_y)$, soit à l'inertie interclasses, en cherchant à travers un algorithme à la maximiser.

L'algorithme de cette technique est comme suit :

Etape 0 : on a la partition $\{\{X_1\}, \dots, \{X_N\}\}$, $k = N$.

Etape 1 : on calcule des distances entre k classes (il y a C_K^2 distances à calculer).

Etape 2 : on agrège les deux classes les plus proches.

Etape 3 : on pose $k \leftarrow k - 1$ on va à l'étape 1.

L'algorithme s'arrête lorsque $k=1$.

Remarque : l'algorithme produit une hiérarchie de N partitions. La 1ère à N classes, la dernière à 1 classe.

ANNEXE II

Tableau27 : La forme de la base de donnée finale de travail pour la régression logistique

	Sol1	Sol2	Rent1			qualité
client1	0	0	1			1
client2	1	0	1			0
client3	0	1	0			1
client4	0	1	1			0
client5	1	0	0			0
.
.
.

Tableau28 : Cordonnées des modalités sur les axes les plus discriminants

modalités	F3	F4	F5	F7	F10	F13	F18	F20	F21
sol_1	0,338	1,724	1,548	0,197	-0,594	0,247	0,772	0,063	0,637
sol_2	-0,303	-0,037	-0,074	-0,301	-0,283	-0,244	-0,047	-0,231	-0,038
sol_3	0,552	-0,424	-0,295	0,588	0,781	0,451	-0,124	0,477	-0,106
rentexp_1	-0,256	0,021	-0,143	-0,002	-0,092	-0,063	0,127	0,076	0,014
rentexp_2	-0,331	0,153	0,264	-0,081	0,115	0,051	-0,041	0,050	-0,291
rentexp_3	1,409	-0,382	-0,167	0,176	0,009	0,065	-0,261	-0,316	0,580
rentfin_1	-0,451	-0,110	0,015	-0,165	0,241	-0,044	-0,205	-0,185	0,130
rentfin_2	0,053	-0,084	-0,158	0,071	-0,020	0,260	0,342	0,367	0,098
rentfin_3	1,026	0,455	0,296	0,266	-0,566	-0,441	-0,208	-0,311	-0,535
serdett_1	-0,196	0,014	-0,047	0,088	-0,047	0,165	0,086	0,001	-0,267
serdett_2	0,128	0,249	-0,256	-0,490	0,106	0,007	-0,084	0,017	0,391
serdett_3	0,132	-0,596	0,681	0,923	-0,139	-0,370	0,006	-0,041	-0,312
frng_1	0,026	0,383	0,462	0,122	0,230	-0,171	-0,252	0,171	0,085
frng_2	0,110	0,227	-0,181	0,318	-0,038	-0,294	0,753	-0,016	-0,133
frng_3	-0,100	-0,601	-0,442	-0,343	-0,253	0,388	-0,164	-0,195	-0,019
chargper_1	0,064	-0,476	0,334	-0,358	0,034	-0,029	0,121	0,078	0,003
chargper_2	-0,229	0,039	-0,551	0,865	-0,381	-0,555	-0,047	0,037	0,173
chargper_3	0,074	1,014	-0,217	-0,025	0,284	0,587	-0,222	-0,208	-0,170
ancint_1	-0,126	-0,583	0,026	0,274	0,002	-0,199	0,131	-0,303	0,256
ancint_2	-0,514	0,375	0,272	0,549	-0,026	0,434	-0,404	0,052	0,034
ancint_3	0,398	0,156	-0,184	-0,518	0,014	-0,135	0,161	0,168	-0,191
qualsec_1	0,331	0,292	-0,340	0,274	-0,687	0,338	-0,148	0,010	0,093
qualsec_2	-0,360	-0,345	0,388	0,233	0,699	-0,326	0,013	-0,170	0,128
qualsec_3	-0,022	0,004	0,008	-0,464	0,086	-0,058	0,134	0,133	-0,199
interet_1	0,037	-0,085	0,497	-0,249	-0,161	-0,208	-0,302	0,318	-0,020
interet_2	0,340	-0,237	-0,876	0,302	-0,030	-0,052	0,390	0,007	0,011
interet_3	-0,228	0,218	-0,003	0,078	0,178	0,238	0,082	-0,322	0,014
partfin_1	0,636	-0,387	0,403	-0,249	-0,110	0,189	-0,092	-0,103	-0,095
partfin_2	-0,385	-0,024	0,003	0,395	0,038	0,276	0,168	-0,252	-0,110
partfin_3	-0,118	0,272	-0,266	-0,143	0,043	-0,337	-0,071	0,263	0,148

PREMIERE PARTIE :

Présentation Générale

montech_1	1,335	0,318	0,400	-0,289	0,956	0,052	0,281	-0,751	-0,128
montech_2	0,009	-0,345	-0,097	0,054	-0,372	0,049	-0,291	0,129	-0,102
montech_3	-0,266	0,158	-0,016	0,022	0,050	-0,042	0,130	0,064	0,090
ratios12_1	0,127	0,306	-0,437	0,066	0,253	-0,385	-0,174	-0,031	-0,080
ratios12_2	0,293	-0,452	0,912	0,186	-0,374	0,183	0,128	0,103	0,256
ratios12_3	-0,357	-0,204	0,140	-0,204	-0,168	0,474	0,189	-0,013	-0,024

Tableau29 : Coefficients (scores) des axes factoriels

axe	score
F1	0,305
F2	-0,546
F3	1,301
F4	-0,880
F5	-1,525
F6	-0,111
F7	2,197
F8	0,062
F9	-0,363
F10	2,471
F11	-0,229
F12	-0,257
F13	1,974
F14	-0,148
F15	0,012
F16	0,777
F17	0,181
F18	-1,517
F19	-0,508
F20	1,328
F21	-1,613
F22	0,240
F23	1,192
F24	-0,593

Tableau30 : Règle de décision (d'affectation)

CLIENT	Fonction
0	1,10834119
1	-1,0324274

Tableau31 : Corrélations canoniques de l'analyse discriminante pour les deux groupes.

Fonction	Valeur propre	% de la variance	% cumulé	Corrélation canonique
1	1,16074631	100	100	0,7329372

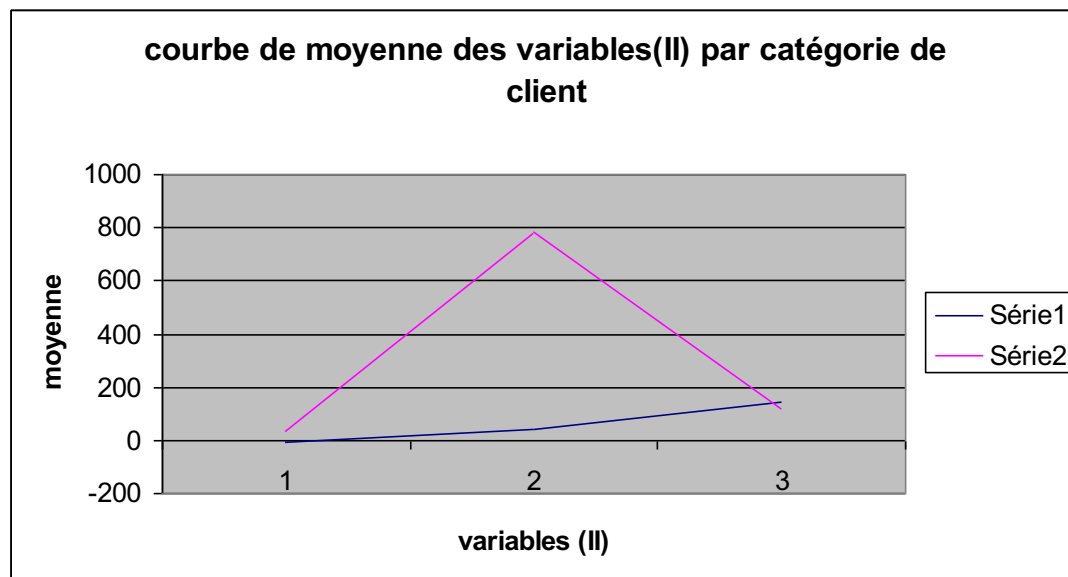
Tableau33 : Les variables I

Variables	code
<i>solvabilité</i>	1
<i>rentexploit</i>	2
<i>rentfin</i>	3
<i>annciété</i>	4
<i>intérêt</i>	5

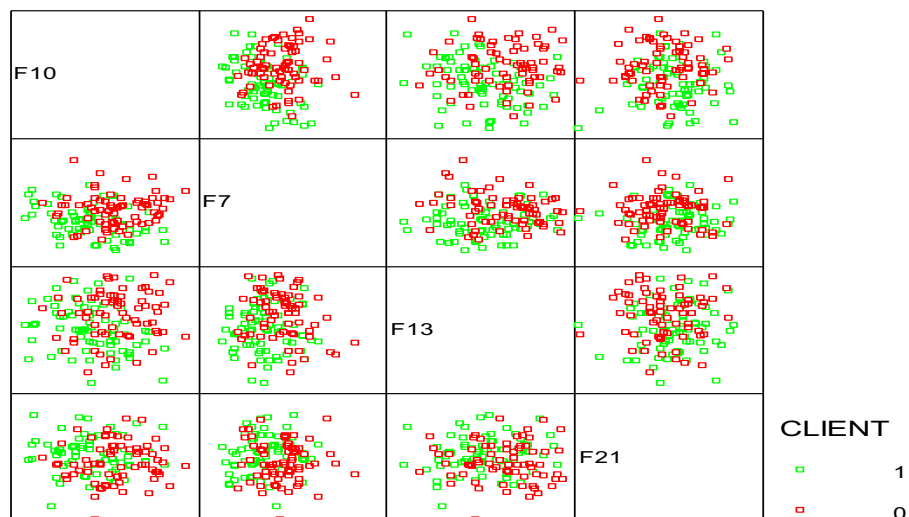
Tableau 34 : Les variables II

Variables	code
<i>FRNG/CA</i>	1
<i>servdette</i>	2
<i>Ratio12</i>	3

Graphique17 : Les moyennes des variables par type de clients



Graphique13 : Dispersion des clients selon F10, F7, F13, F21



Graphique14 : Processus de suivi et d'octroi de crédit

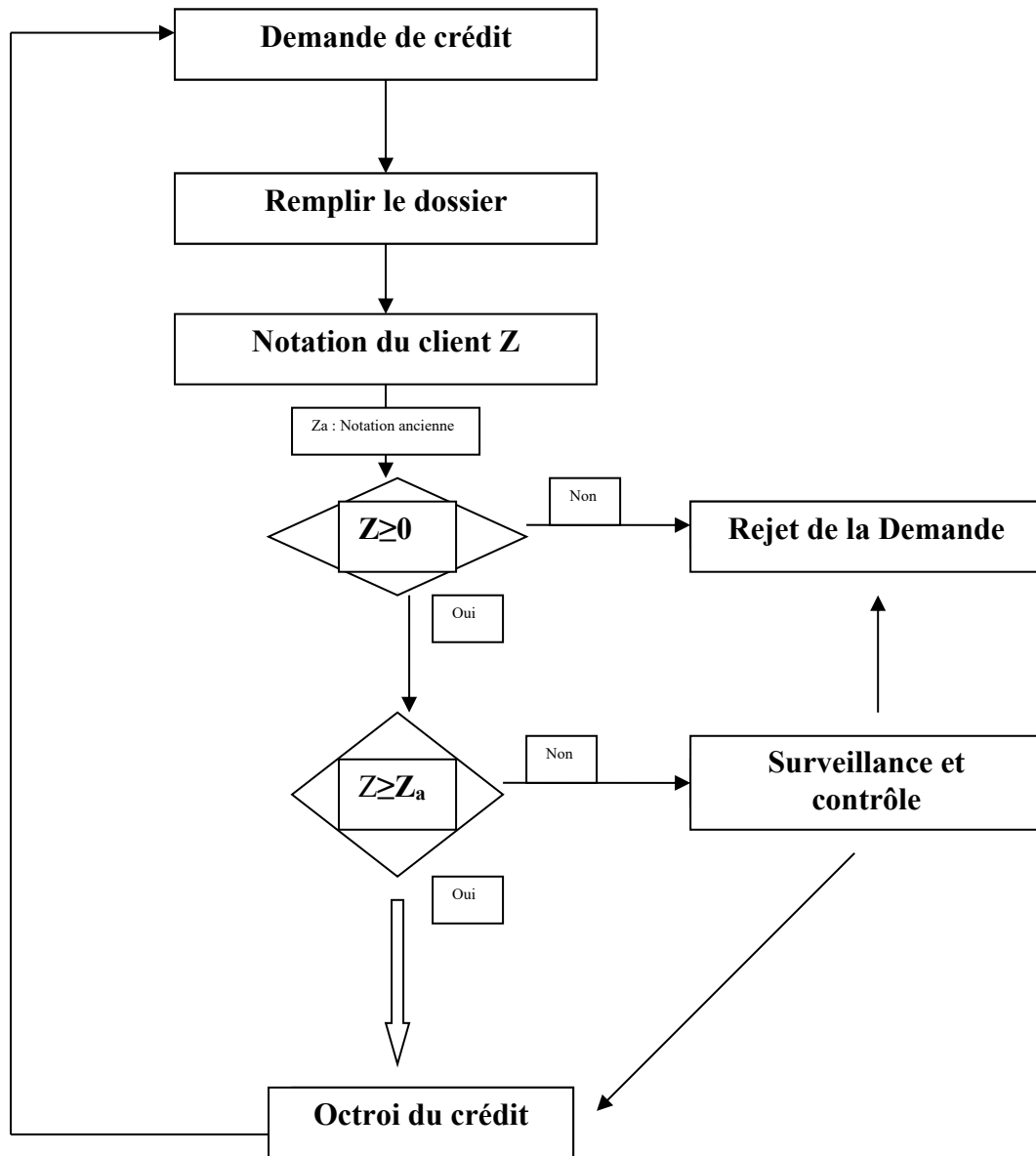


Tableau 32 : Quelques critères de comparaison entre les méthodes

	ADL Fisher	ADQ Fisher	AFD	Régression logistique	Arbre de segment ation	Disqual	Réseaux de neurone s
Méthode descriptive	D	D	O	D	D	D	
Méthode décisionnelle	O	O	N	O	O	O	O
Méthode paramétriqu e	O	O		O	N	N	N
Outil sensible à l'échantillon				O			
Outil dépend de la taille du groupe	N	N		O			
Outil dépend des seuils	N	N		D*	O	D*	O
Méthode robuste aux valeurs extrêmes	N	N		D**	O	O	O
Interprétabili té facile	O	N		O	O	O	N
Possibilité d'utiliser des variables qualitatives	D	N	N	O	O	O	O

Légende : O : oui N : non D : cela dépend

* Des seuils interviennent si les variables quantitatives sont rendues qualitatives

** Cela dépend si les variables utilisées sont continues

Source : « Analyse discriminante : application au risque et scoring financier »

Figure15 : Architecture de la réforme Bale II

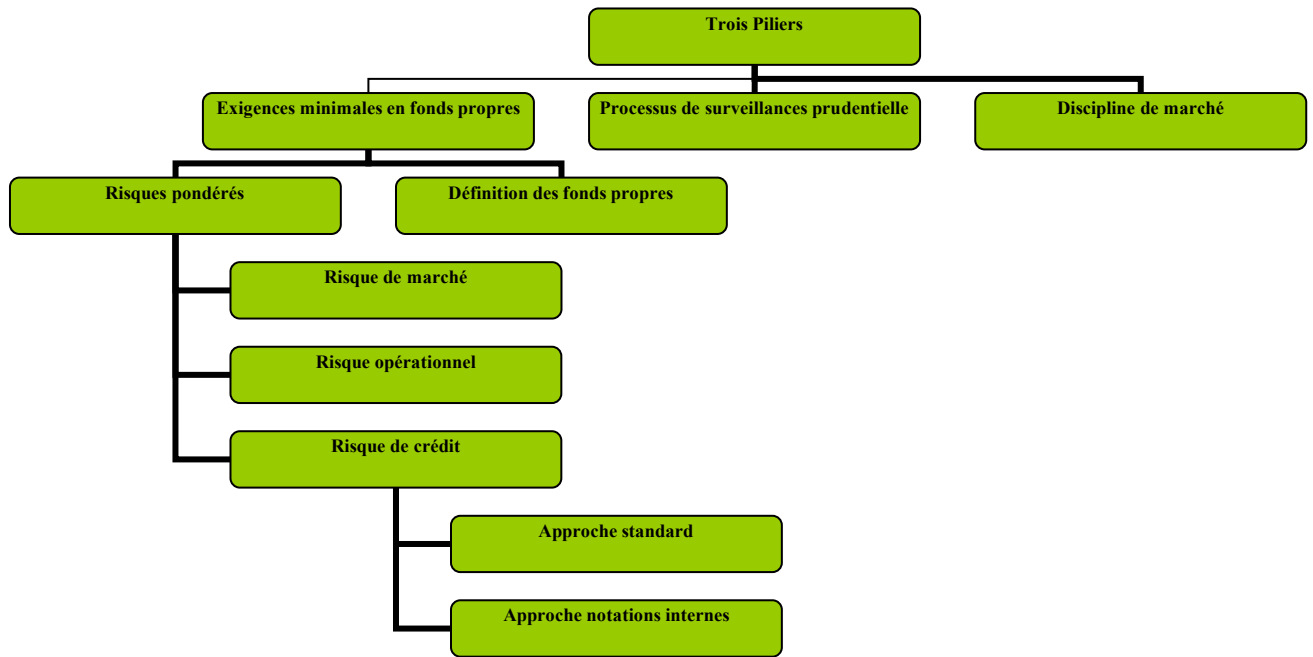
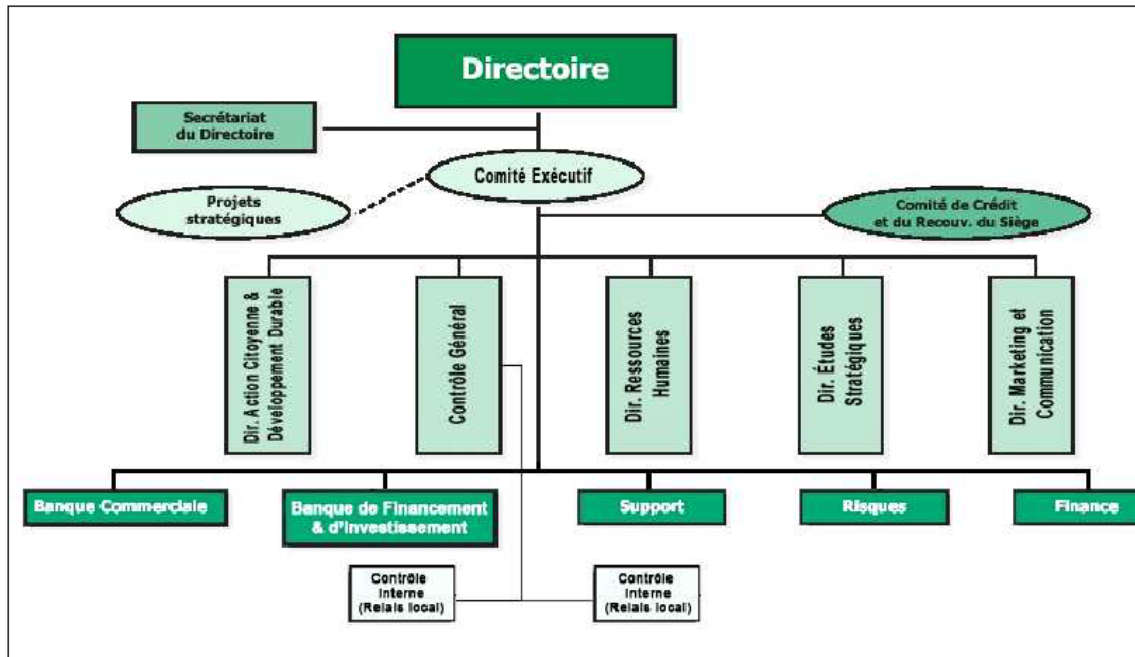


Figure 1 :Organigramme



Bibliographie

Ouvrages

- **Mireille.B (2001)** « Analyse discriminante : application au risque et scoring financier»
- **Eric. M (1990)** « Analyse bancaire de l'entreprise. Méthodologie »
- **Michel.J (1999)** « Méthodes de base de l'analyse des données »
- **Banque de France** « Détection précoce des défaillances d'entreprises dans le secteur de la construction »

Publications

- **Comité de Bale sur le contrôle bancaire (Juin 2004)** « Convergence internationale de la mesure et des normes de fonds propres », Banque des Règlements Internationaux.
- **Projet de fin d'étude (juin 2005)** « Elaboration d'un modèle statistique d'évaluation de risque de contrepartie au sein de CCG »

Notes de cours

- **Mr. Nsiri Said Ramadan (2007)**, cours d'analyse de données, INSEA.
- **Mr. Touhami abdelkhalake (2007)**, cours d'économétrie avancée, INSEA.
- **Mr. Benchekroun kamal (2007)**, cours d'analyse de la variance, INSEA.

Sites Web

- www.bankelmaghreb.com
- www.cedric.cnam.fr